

Otto von Guericke University Magdeburg

Faculty of Computer Science



Bachelor Thesis  
Computational Visualistics

# Parallel Coordinates for Narrative Visualization

*to communicate risk factors in cohort data to a broad audience*

Author:

Beatrice Budich

June 9, 2023

Advisers:

*Prof. Dr.-Ing. habil.* Bernhard Preim

*Dr.-Ing. habil.* Monique Meuschke

Department of Simulation and Graphics

**Beatrice Budich:**

***Parallel Coordinates for Narrative Visualization***

*to communicate risk factors in cohort data to a broad audience.*

Bachelor Thesis, Otto von Guericke University

Magdeburg, 2023





## ***Abstract***

Large epidemiological population studies collect data on a cohort to identify risk factors for various diseases. The aim of a communication tool can be to convey preventable risk factors to a broad audience. In this work, I develop and test the possibilities, techniques and processes of a narrative approach. Multivariate time-dependent cohort data are presented with parallel coordinates embedded in a narrative setting. AI-generated protagonists, interactive incremental chart construction, annotation, gamification elements and cognitive behavioural therapy influences are part of the web-based story design. In a small case study, subjects were tested with the think-aloud method for qualitative feedback, as well as with a questionnaire on memorability, understanding of the parallel coordinate diagrams and user engagement. The results were satisfactory and showed a positive correlation between focus attention and memorability. The results were independent of demographic characteristics.

***Keywords:*** *External Science Communication, Narrative Visualization, Parallel Coordinates, Cohort/Empirical Data, AI Character Design*



# Contents

<b>1</b>	<b>Introduction &amp; Motivation .....</b>	<b>9</b>
<b>2</b>	<b>External Science Communication .....</b>	<b>15</b>
	2.1 Heterogeneity of a Broad Audience .....	15
	2.2 User Engagement & Memorability .....	17
	2.3 Narrative Visualization .....	19
	2.4 Visualizing Cohort Data .....	25
	2.4.1 Visualizing Multivariate Data with Parallel Coordinates ...	26
	2.4.2 Time-oriented Parallel Coordinates (Temporal PC) .....	31
	2.5 Additional Concepts: Gamification & Cognitive Behavioral Therapy .....	34
<b>3</b>	<b>Data .....</b>	<b>37</b>
	3.1 SHIP Data .....	37
	3.2 Description of Selected Variables .....	38
	3.3 Data Preprocessing, Treating Missing Values & Outliers ...	45
	3.4 How to Find a Protagonist in Data? .....	47
<b>4</b>	<b>Story Design .....</b>	<b>53</b>
	4.1 Target Group Characteristics .....	53
	4.2 Design of Protagonists .....	55

4.3 AI-generated Character Design .....	56
4.4 Story Design .....	75

## **5 Implementation ..... 86**

5.1 Web-based Concept .....	86
5.2 Documentation of Methods .....	86

## **6 Evaluation ..... 87**

6.1 Early Feedback .....	87
6.2 User Study .....	87

## **7 Results & Discussion ..... 91**

## **8 Conclusion & Future Work ..... 93**

### **I. Bibliography ..... 97**

### **II. Appendix ..... 103**

- Data
- Questionnaire
- Sketches

### **III. Statutory Declaration ..... 119**



# 1 Introduction & Motivation

**Motivation and Relevance:** Large epidemiological population studies, such as the Study of Health in Pomerania (SHIP), collect data about participants to screen the risk factors for various diseases, such as obesity, diabetes and liver diseases [Völzke 2011]. Epidemiology is concerned with the prevalence, causes, and consequences of health-related conditions and events. This information is used not only to determine risk factors for specific diseases, but also to “inform people about healthier lifestyles” [Klemm 2015]. The message of a *communication tool* might be to promote prevention, i.e., convey *avoidable risk factors* as well as the importance of a healthier lifestyle. The intent must be adapted for a specific health topic in collaboration with the domain experts.

**Goal:** With this work, I aim to explore opportunities, techniques, and the process of creating a communication tool for a broad audience to convey medical cohort data and knowledge about preventable risk factors. A cohort is a “set of individuals sharing a commonality as a group” [Withers 2009], e.g., age, experiencing the same historical event, geographical boundaries, and who then age together [Rosow 1978].

A cohort data set contains multivariate and time-oriented data. In the visualization community, *Parallel Coordinates* (PC) are well known for its strength in displaying *multivariate data*, i.e., presenting an *overview of clusters and trends* as well as a fast recognition of *outliers* [Palmas 2014]. This advanced visualization is typically used by data analysis experts in specific domains. However, a few authors mentioned the potential of Parallel Coordinates for communicating data to a broad audience [Siirtola 2009, Johansson & Forsell 2016].

On the other hand, some authors warn that the representation can lead to an overload of information due to clutter of overlapping lines. The amount of clutter

depends on the amount of data, namely number of records but also on the techniques to reduce cluttering, e.g., edge bundling, use of transparency and interaction techniques [Inselberg 2008, Heinrich & Weiskopf 2013, Palmas 2014]. I will discuss these issues and explore ways in which *multivariate data* as well as *time-oriented data* can be integrated into PC as well as how these techniques can be adapted to make them understandable and meaningful to non-experts.

For this purpose, I explore techniques of *Narrative Visualization* – as a combination of Storytelling and Visualization – and other ‘access techniques’ to introduce Parallel Coordinates to a lay audience as well as reduce complexity. Moreover, techniques such as ‘incremental chart construction’ and ‘personalization of diagrams’ help to increase *User Engagement* and *Memorability*.

**Lack/Knowledge Gaps:** There is a lack in the literature about Parallel Coordinates for Narrative Visualization. None of the papers directly address the research questions of ‘How do you technically optimize Parallel Coordinates for Narrative Visualization?’ or ‘What strategies can be used to adapt parallel coordinates for the general public?’ In addition, because cohort data have a temporal dimension, the question arises: ‘How time-oriented data can be represented with this visualization type in the context of Narrative Visualization?’

Johansson & Forsell (2016) mentioned that “new enhancements of Parallel Coordinates [...] would be of interest for a wider audience” and “although users tend to be confused at the beginning they quickly learn how to use parallel coordinates” and appreciate to interact with the visualization. Therefore, with a little support Parallel Coordinates could work for a broader audience.

Weitz (2020) suggests Parallel Coordinates could be used in the field of Storytelling, but warns that “Parallel Coordinates Plots are usually difficult to understand for non-technical audiences” [Weitz 2020], especially when over-plotting and cluttering occur.

To bridge the gap between the expert’s field and the requirements of a broad audience especially a general public, I will provide a literature review of techniques for time-oriented Parallel Coordinates and discuss their level of complexity, advantages and disadvantages with respect to the target audience.

**External Science Communication:** For a public debate on scientific data, there is a need for adapting data visualization with the goal that a lay audience can understand and draw conclusions on data. To narrow down the target audience, we look at the *general public*, which is heterogeneous in respect to knowledge and interests, age, educational backgrounds, abilities and expectations; has no expert knowledge in any particular scientific field, nor in visual data

analysis. The intended audience is a subset of the general public who, while not inherently motivated to explore data, but is generally interested, open, and motivated to explore and learn – up to a certain level of information complexity [Böttinger 2020].

**Data:** The story design and data visualization are based on the SHIP data from 1997–2012 with three time steps having an average temporal distance of five years between the records [Ittermann 2012]. However, the concept can be adapted to related data and specific health problems, such as back pain. For this purpose, I have documented the reasoning for my design choices in terms of extension or generalization.

## Main questions

This thesis addresses the following main questions:

The overarching question is: How can information from *cohort data* be communicated to a *broad audience* and what are needs of this target group?

**Q1:** How can *Parallel Coordinates* be technically optimized to visualize *multivariate data* to a broad audience?

**Q2:** How can *Parallel Coordinates* be modified to visualize *time-oriented data* to a broad audience?

**Q3:** How to design a *data-driven story* and its *characters* based on *cohort data*?

## Main Hypotheses

**H1:** Parallel Coordinates – as a suitable visualization type for multivariate data – has the potential to be used in data stories for the general public (External Science Communication). To exploit the potential of Parallel Coordinates in terms of *perceptions*, *comprehension* and *memorability* of cohort data to proof or provide a message, techniques for *introducing and modifying* the visualization are needed.

In this thesis I investigate opportunities and ‘access-techniques’ (aids) to make Parallel Coordinates and its data accessible for a broader audience

- *A Metaphoric Introductory Example* to familiarize the user with the type of visualization by having the user solve simple tasks with a slightly increasing level of difficulty.
- *Incremental Construction* to prevent the user from an overload of information by a user-controlled interaction pace (step-by-step processing). This technique supports different user-dependent requirements regarding the time needed to process the information.

This step-by-step concept is integrated throughout the story and slowly increase the overall complexity of a diagram – namely in number of axis, data records and displayed time steps, for example the first PC only displays one record in one time-step. An optional ‘backward button’ offers the possibility to restart the diagram construction process.

- *Icons/Pictographs* to support comprehension and memorizing of a variable through an easily recognizable and memorable image (personalization) and provide visual orientation within the layout.
- *Annotations* are designed to be short, clear and meaningful and provide comprehension. Longer explanations are supported by a tooltip format that can be accessed via an icon. This serves to reduce complexity (text-based and auditory).
- *Storytelling* techniques to embed the Parallel Coordinates in a meaningful context and for enhancing user engagement and memorization. To support understanding a doctor protagonist introduces the diagram and draws a conclusion and emphasize the most important information.
- In addition, I integrated *Interactive/Gamification Elements* that provide introducing Parallel Coordinates for example.

**H2:** Smaller subsets of the data are needed to prevent the user from an overload of information. The first step would be to display only one participant. I developed two types of incremental construction of time-oriented PCs and evaluated the test persons’ preference for the variants. In the future work part, I will discuss further opportunities to show clusters in time-oriented data.

**H3:** To communicate *cohort data*, the data visualizations must be embedded in a *meaningful* context. To do this, I use a data-driven storyline with a ‘focus-context concept’. Focus is a single subject from the data set which serves as an *anchor* for the user in the context of the cohort data. The principle works best with increasing identification with the protagonist so that the goal of character design is to create a character that is as sympathetic and vivid as possible.

I used a textual description based on the data for the selected subject – to generate semirealistic images for the protagonists by AI image generation and animate this image with AI video generation. A single image is taken as starting point or ‘seed’ from which variations in terms of emotional expression, age and weight can be generated by image-to-image translation or machine learning by using the same name and a seed image as well. By combination of image and text a video also can be generated with AI support. The combination of interaction and video sequences increases variety of the story design and serves different preferences among users.

**Prototype:** My main theses will flow into a concept and implementation of a web-based prototype. The website contains a storyline with AI generated images and video sequences of the protagonists, auditory dialogs, interactive and annotated data and information visualizations (Parallel Coordinates, Isotype charts and Pie Charts), interactive 3D models of different health stages of liver, gameplay and other interactive elements, a start menu and navigation elements are also supported.

**Evaluation:** The story is about risk factors of *non-alcoholic fatty liver* and the recovery of a female patient-protagonist accompanied by a male doctor. I evaluate the impact of the user engagement techniques and design choices on *comprehension*, *memorization* and *user engagement* with a focus on the Parallel Coordinates plot integrated into an interactive web-based prototype.

I test variants of the data visualization that differ in complexity within the same story framework. At the beginning of a test, I ask the test persons if they are familiar with data visualization and in particular with Parallel Coordinates to examine the impact of the ‘access techniques’, e.g., introductory example. In addition, I regularly collected *early feedback* from potential users during the prototype development.

The story will contain tasks. The time that a person needs to walk through the story is also measured and compared with the statement if the story was felt too long. To get *qualitative feedback* the users were asked to *think aloud* during the story exploration and asked some open questions at the end. After this participants filled out a questionnaire to evaluate comprehension, memorization and user engagement.



## 2 External Science Communication

### 2.1 Heterogeneity of a Broad Audience

Creating data visualisations for a wide audience includes many requirements and challenges that need to be considered during the design process.

The broad audience consists of several subgroups which differ mainly in the amount of domain knowledge. The largest subgroup is the ‘general public’. This group is heterogeneous “in knowledge and interests, varying age groups, varying cultural, geographical, and educational backgrounds” [Böttinger 2020].

Böttinger et al. (2020) give a slightly *narrower definition* in the context of creating visualizations for the general public. Those ‘recipients’ (of visualizations) from which we

- (1) cannot expect prior expert *knowledge* of the underlying science but can have previous or related knowledge, and
- (2) we cannot expect experience with visual *data analyses* in general.
- (3) Furthermore, the recipients have some portion of “*openness and interest*” to understand the main messages of the data presented” is assumed [Böttinger 2020].

To formulate a clear *intent* for the visualization or story is the starting point and defines the purpose of the visualization in the context of the target groups.

“Failure to understand the needs of the audience can lead to misinterpretation of the data and mistrust of the information” [Böttinger 2020]. To reduce this risk, we need to take a closer look at demographic and psychographic characteristics of the target audience, which are important aspects in the analysis phase of a design process.

*Demographic* properties are, e.g., gender, age, occupation, and level of education. *Psychographic* aspects embrace sociological and emotional factors, such as interests, needs, expectations, preferences (e.g., between a video presentation and an interactive application and skill level (here: using website or interactive products) [Böttinger 2020], as well as disabilities, recognition abilities and memory capacity. In chapter 4, I describe these properties of the intended target group in more detail.

Böttinger et al. (2020) state that advanced visualization can be open-up to a broad audience. On the basis that there is general interest, openness and motivation to explore and learn, and since Parallel Coordinates are an advanced visualization, there have to be an introductory example.

## Complexity Reduction

*“Sometimes reality is too complex. Stories give it form.”* (Jean-Luc Godard)

We must keep in mind that a non-expert is probably not willing to deal with too much complexity. For this reason, Dykes (2019) suggests using storytelling to reduce complexity, and Böttinger et al. (2020) explain that data visualizations require a clean, simplified visualization design. Specifically, they suggest to clean the visualizations and displays from anything that might conflict with *communicating a clear message*. But there has to be a compromise between *scientific “correctness, completeness* and the quest for a *simplified visualization design.*” [Böttinger 2020] Here, a leading question for the information reduction could be: ‘What information or detail is needed to communicate the message or tell the story?’

Aspects of *visual design* have a major impact on the level of cognitive load. For example, colormap design is an important facet in information reduction with respect to a broad audience [Böttinger 2020]. The general advice is to use a small set of colors with respect to hues where each color encodes a different meaning, to use mostly unsaturated, pastel-colored tones and transparency mapping for the context information and background. Saturated colors should be only used for highlighting and to emphasize the most important information [Wang 2008, Böttinger 2020].

Wang et al. (2008) summarized a useful set of rules for color design of illustrative visualizations and have shown that colors are better to discriminate when they differ in each dimension of hue, saturation and brightness. Furthermore, layout structure and usage of typography also play an important role in information reduction. Not only does the information need to be presented in a structured way, but there is also the question of how to guide the user through the information [Böttinger 2020].

In summary, the most important aspects a broader audience may expect are: clearness regarding the information and layout arrangements, comprehensiveness and support, accuracy to a certain degree, reduced and simplified information, a higher proportion of illustration (images, info graphics etc.) than textual information and to avoid the feeling that the story ‘is too long’.



## 2.2 User Engagement & Memorability

An important commonality, is that users of a broader audience are most likely not inherently enthusiastic about the underlying science or data visualization. Therefore, appropriate techniques are needed to engage users, such as storytelling, a creative visualization design, user interaction, or other means [Böttinger 2020].

«What makes a visualization memorable?» is one of the most interesting research question in the field of data and information visualization and is closely related to *User Engagement*. A user who feels *cognitively and emotionally involved* would more likely remember the information he perceived [Borkin 2013].

«What is (User) Engagement?»

O’Brien and Toms (2007) define engagement as a *quality of user experience* characterized by the listed *attributes* below. I used the User Engagement attributes as categories for the techniques I collected from literature:

1. Aesthetic and Sensory Appeal
  - Personalization of charts, e.g., isotype chart [Kosara 2016],
  - Design enhancements (even simple decorations) [Peña 2020].
2. Attention
  - Uniqueness, symbols (icons & pictographs) [Kosara 2016, Bateman 2010],
  - Unusual techniques that stand out [Kosara 2016, Bateman 2010],  
e.g., working with AI generated material.
3. Awareness
  - Storytelling / Identification with characters or topic.
4. Feeling of being-in-control
  - Incremental Construction, Step-by-step building interaction [Kosara 2016].
5. Interactivity
  - Flow and fluid interaction [Amini 2018],
  - Incremental Construction [Kosara 2016],
  - Tooltips,
  - Interacting with a 3D model,
  - Navigation, Exploration, Interactive tasks,
  - Gamification, e.g., Interactive tasks, getting points and honor,  
the user can push the story forward by interaction.

6. Novelty and Variety
  - Introductory example [Böttinger 2020],
  - Variety is provided through various (interaction) techniques, visual design elements and media,
  - The combination of all the techniques together I used in the data-driven story prototype is very new to users.
7. Challenge
  - Tasks,
  - Increasing complexity of Parallel Coordinates.
8. Feedback
  - Either positive feedback or neutral feedback,
  - Implemented either as response to a click- or hover-event.
9. Interest
  - A general interest and openness to health-related topics is assumed after the definition of the target group [Böttinger 2020].
10. Positive Affect / Pleasure
  - Satisfying and pleasurable emotions,
  - Positive user experience, captivated and motivated,
  - Flow, presence, immersion, enjoyment, and playfulness [Amini 2018],
  - Sense of humor.

*Design enhancements* (even simple decorations), if relevant to the data, can improve recall of title and thematic elements of visual displays [Peña 2020] which is an example showing the close relationship between engagement and memorization.

*Narration* is a technique to reduce complexity and promote memorability of content such as messages and facts. Many of the user engagement attributes (listed above) can be embedded in the concept of Narrative Visualization, which we consider in more detail in the next section.

## 2.3 Narrative Visualization

Many studies have shown that information embedded in a narrative context is more likely to be *memorized*. Mar et al. (2021) found in a meta-analysis about memory and comprehension of narrative versus expository texts with more than 33,000 subjects that “stories were more easily understood and better recalled than essays.”

Thus, *narration* has a high relevance in the context of a *broader audience*. “Whether the data visualization takes the form of an infographic, an inter-active visualization or a guided visual presentation” [Böttinger 2020] it must either be presented within a *narrative structure* that captures the audience’s interest. However, a visualization itself can also have a narrative structure, such as an annotated chart.

*Narrative Visualization*, a combination of *Storytelling* and *interactive graphics* [Segel & Heer 2010], yields as an effective technique to engage a broad audience create improved learning experiences [Kleinau 2022]. In this context, visualization has to be understood as a supportive element within the narration/ messaging.

As Mittenentzwei et al. (2023) note, particularly for online media, there is a *trend* toward increasingly *advanced interaction techniques*.

### Storytelling

Storytelling is the strategy of guiding the user through a topic by a chain of events, highlighting, view direction elements such as arrows instead of overwhelming the recipient with facts, messages and information [Kosara 2013].

### Story Components

Dykes (2019) name three essential components of data-stories: data, narrative, and visuals. A further approach is to consider four components: (1) content, (2) characters, (3) conflict, (4) structure [Mittenentzwei 2023].

First, the *content* of the story that is based on the data and additional information, e.g., facts received from domain experts.

Second, *Narrative Characters* are designed to convey the message of the story, the aim of the character design is that the user is be able to sympathize and identify with them. The difference to classic media with a presenter or narrator is that storytelling based on data visualization requires an alternative form of a

narrator. For example, protagonists tell their story. The representation has to consider a suitable personality, visual appearance and behavior characteristics for that role.

Third, a *Conflict Situation* communicates an issue which should be resolved. In medical cohort studies this would be a widespread disease or health problems, such as non-alcoholic fatty liver (approx. one-third of adults [Ittermann 2022]) or back pain (it is estimated that four out of five experience back pain during their life [Roland 2007]).

Fourth, a story needs a *narrative structure*. There are various approaches that look at this topic from different perspectives.

## Narrative Structures

*Freytag's Pyramid* is a narrative structure originally came from Ancient Greek theatre is one of the most popular narrative structures which provide a guided representation. The story can be divided into six parts:

(1) *introduction*, (2) *incidence*, (3) *rising action*, (4) *climax*, (5) *falling action*, (6) *resolution* [Yang 2021].

While a story with a Freytag's Pyramid structure classically ends with a catastrophe in which the main character is ultimately undone by their own choices and actions, *Campbell's Hero Story* [Campbell 2008] is characterized by its strong focus on the successful development of the protagonist.

Furthermore, a narrative structure can be restricted to moving forward and backward in a chain of story segments called a *linear structure*. A more *elastic structure*, on the other hand, allows the user more freedom of exploring different optional paths. However, there is a main thread that contains the most important information [Yang 2021].

Segel & Heer (2010) note that an important strategy of narrative visualization is to find a balance between the *author-driven* narrative flow and *reader-driven* part. *Reader-driven* aims at the reader's discovery of the story. This mental processing would have a significant impact on understanding and remembering the content of the story. Important techniques here could be quizzing, gamification elements, user-driven incremental construction and other *decision-making* opportunities.

*Author-driven* elements on the other hand provide the user in terms of a supportive structure or that there are restricted paths to guarantee that a user gets presented the most important information

Images with Copyright may not be displayed

*Figure 1: Narrative Structures with different balance between author-driven and reader-driven portion after Heer & Segel (2010), adapted by M. Stefaner (2012)*

Segel & Heer (2010) describe the ‘*Drill-Down Story*’ as a structure with a reader-driven design where the user can choose different paths of a topic. ‘*The Martini Glass*’ has a tight narrative author-driven path at the beginning of the story that opens up later in the story in respect of free exploration to the reader-driven part. Therefore, at the end of the story this structure requires a lot of independent exploration and processing of the information on the side of the user. This structure may be effective when users have some founded background knowledge or visualization experience. However, the open-end structure may be confusing for users that are new to topic or other novel elements, such as advanced data visualization types. The broad audience probably needs a clear end of the story and a carefully guided representation [Böttinger 2020].

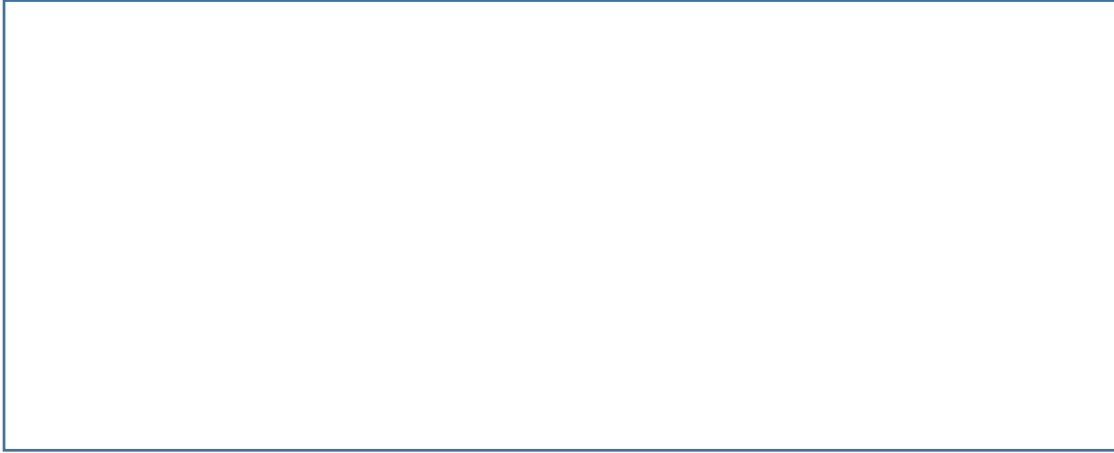
‘*Interactive Slideshow*’ on the other side has a more author-driven structure with reader-driven elements. “Interactive slideshows work well with both complex datasets and narratives” [Segel & Heer 2010]. This approach based on a format a broad audience is familiar with, from the usage of websites or games for example. Segel & Heer (2010) note that there is a more balanced mix of author-driven (messaging) and the reader-driven (interactivity) in this narrative format. Therefore, the interactive slideshow structure is a good starting point for an application for a broader audience.

Therefore, an ‘*Interactive Slideshow*’ is the most linear structure, a ‘*Drill-Down Story*’ is the most elastic structure and ‘*The Martini Glass*’ is a balanced mixed structure – linear at the beginning and turns later into an elastic structure.

## **Narrative Genres**

These mentioned elements of Narrative Visualization could be arranged in different genres according to Segel & Heer (2010), who introduced the following seven genres:

- (1) Magazine Style, (2) Annotated Graph/Map, (3) Partitioned Poster,  
(4) Flow Chart, (5) Comic Strip, (6) Slide Show, (7) Film/Video/Animation.



*Figure 2: Seven Genres of Narrative Visualization after Heer & Segel (2010), adapted by M. Stefaner (2012), p. 23.*

## **Narrative Medical Visualization**

There is little research about the combination of medical visualization and narration [Meuschke 2021]. Mittenentzwei et al. (2022) and Kleinau et al. (2022) present workflows for creating medical data-stories and evaluated them in user studies.

In medicine Narrative has the potential to improve “patient understanding of diagnostic procedures and treatment options, promote confidence, reduce anxiety, and support informed decision-making” [Kleinau 2022].

To communicate risk factors for ‘high blood pressure’ and ‘cerebral small vessel disease’, Mittenentzwei et al. (2023) adapt the *Campbell’s Hero’s Journey* for narrative medical visualization to a *patient-centered story*.

The storyline can be divided as follows:

- (1) Hook, (2) Patient’s life, (3) Awareness/Relevance,
- (4) Meeting the Physician, (5) Diagnostics/Diagnosis,
- (6) Treatment, (7) Lifestyle adjustment, and (8) Adherence.

This storyline can also be mapped onto the Freytag’s Pyramid, with a climax in the diagnosis section. Here, for example, warnings of consequential diseases could be given to create a peak in the tension arc. Furthermore, the final could rather be designed positively, describing the success of a patient who makes lasting changes to his or her lifestyle.

This structure can also be modified, for example, the patient’s lifestyle can be represented even after diagnosis. This structure generalizes well to other medical topics.

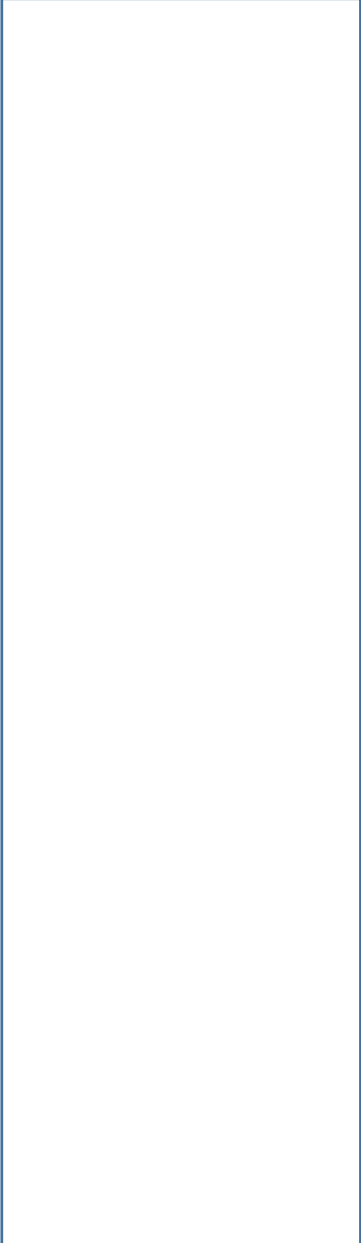


*Figure 3: Adapted Freytag’s Pyramid for communicating risk factors for health problems [Mittenentzwei 2022]*

Kleinau et al. (2022) worked with a similar narrative structure for ‘aortic blood flow vortices’ and compared a ‘*Slideshow-based*’ with a ‘*Scrollytelling-based*’ navigation technique. Their case study found that users preferred the click technique of the slideshow format. Some users had trouble recognizing scrolling as navigation technique, and many participants found the scrolling speed too slow. This can be interpreted to mean that the match between the mechanical movements (input) and the visual feedback on the display was not well matching.

However, scrolling is widely used in online media. So if the match between input and display is appropriate, scrolling has potential for storytelling.

They also mentioned an important principle regarding user orientation within a layout. Users need an arrangement of the layout elements in ‘reading direction’, otherwise they will be confused. Other arrangements need to be justified by a strong visual hierarchy, emphasizing the first element to which the users should direct their attention.



Another narrative approach to communicate medical data with an additional psychological and social-economic context is proposed by So et al. (2021). The authors developed a narrative structure using a ‘*layer technique*’ based on the *martini glass structure*.

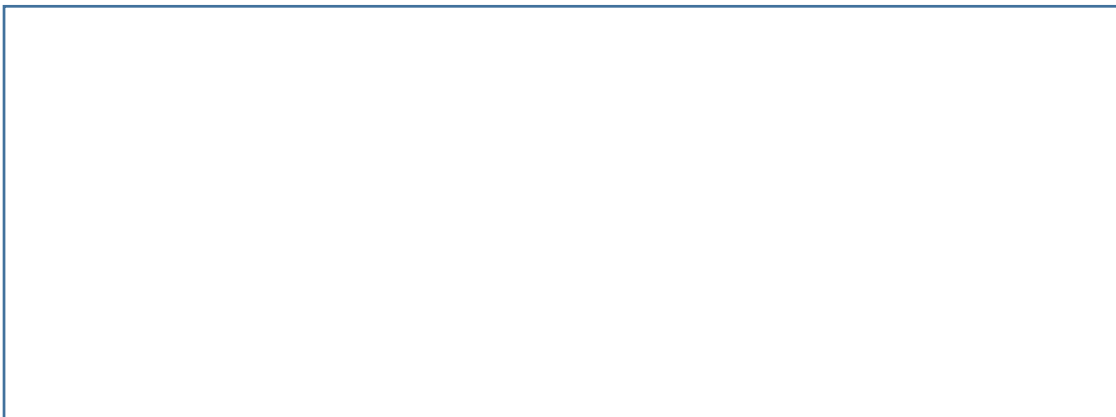
They describe *four layers* that represent different domains involved in a health problem:

- (1) biological layer, (2) bio-psychological layer,
- (3) psychological layer, and (4) social layer.

These layers contain different types of visualizations, i.e., a *bubble visualization* for symptoms, *annotated anatomical 3D model* highlighting affected body parts, and cartographical *maps* showing the distribution in different areas for the social layer.

The layers can be explored by *scrolly-telling technique*. In user case study they found that their approach was effective for “user awareness but also in promoting opinion change” [So 2021].

→ *Figure 4: The storyline of the visualization [So 2021].*



*Figure 5: Visualizations for different aspects in medical data [So 2021].*



## 2.4 Visualizing Cohort Data

Since cohort data are both multivariate and time-oriented, this part of my research relates to the question of which data visualizations are appropriate for this purpose. I will first focus on *multi-variate data visualization* followed by a brief overview of a classification for *time-oriented data visualizations*.

In the final step, I will merge multivariate and time-oriented properties into a literature review about *time-oriented Parallel Coordinates* and discuss which solutions are suitable for communicating medical data to a broad audience.

Guiding questions for the discussion:

- (a) What aspects make a data visualization suitable/adaptable for a broad audience?
- (b) Which features should be avoided with respect to information overload (or support complexity reduction)?
- (c) What are limitations and how adaptive is the visualization type in terms of scalability?

First of all, experts must conduct, evaluate and carefully interpret studies to validly determine risk factors. Then they can be used for narrative purposes.

### 1. Visualizing risk factors

Goal: Show correlations between certain risk factors and occurrence of a disease or health problem by a clear distinction or differentiation between participants who have a health problem from healthy participants.

### 2. Visualizing diagnosed-related variables

Goal: In diagnose certain values for example from a blood test show the health status, i.e., the expression of a health problem. For example, the liver enzymes GGT, ASAT and ALAT are used to determine if a fatty liver is inflamed or damaged [Ceriotti 2010].

As we have noted, ‘complexity reduction’ is important for a broad audience. First, there is the question of using two- or three-dimensional visualization. Due to the problems with projection and occlusion in three-dimensional visualizations, two-dimensional data visualizations are more popular especially in the area of information visualization addressing general public. Moreover, the spatial imagination is varying between individuals or is considered as a skill that has to be developed through training [Górska 2005, Norboy o’g’li 2020].

Wegman (1990) found the classic scatterplot is a fundamental tool in data analysis to detect clusters and outliers but is very limited by a maximum of three dimensions.

A systematic overview of multidimensional/multivariate data visualization is provided by Chan (2006): The author first classified visual data exploration techniques into six categories namely: (1) geometric, (2) icon-based, (3) pixel-oriented, (4) hierarchical, (5) graph-based and (6) hybrid techniques.

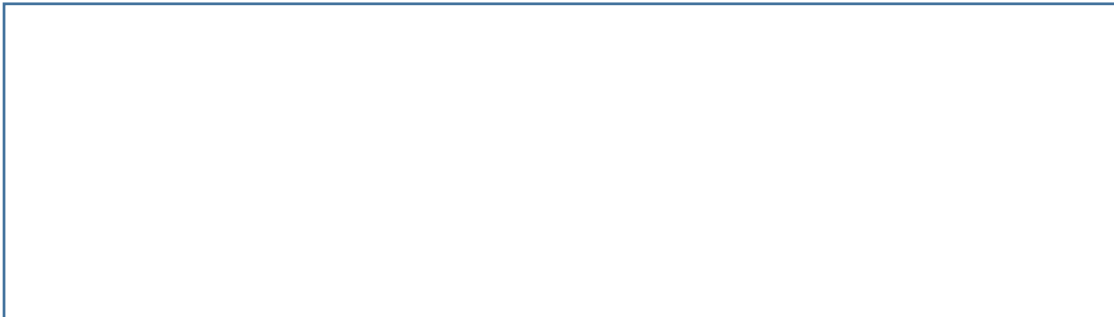
Categories (2)–(6) describe in more detail how they represent the data geometrically, while category (1) contains representations that do not fit into the other categories. Classes (2) and (3) are named after the graphical objects they use to represent the data, such as icons or pixels. Hierarchical and graphics-based techniques describe structures.

Regarding geometric projection attributes can be mapped to a “typical Cartesian plane like scatterplot, or more innovatively to an arbitrary space such as parallel coordinates.” [Chan 2006]

The author provides an overview about Multivariate data visualization types, such as Scatterplot Matrix, Projection Matrix, HyperSlice, Hyperbox, Parallel Coordinates, Andrews Curve, Radical Coordinates Visualization, Star Coordinates, Table Lens.

#### **2.4.1 Visualizing Multivariate Data with Parallel Coordinates**

Parallel Coordinates are a visualization type that is suitable and established for multivariate data in different applications and variations. In 1990, Inselberg and Dimsdale pioneered the use of Parallel Coordinates in the Information Visualization community. Heinrichs’ book about Parallel Coordinates (2013) gives an illustrative overview about the state of the art of this visualization type.



*Figure 6: Duality relation between the points in 2D Cartesian space and the lines of a PC chart [Heinrichs 2013, p. 19]*

This visualization technique contains parallel axes where each axis stands for a single dimension or variable. The layout can be horizontal or vertical, however the horizontal layout is being most common used.

An entity, e.g. a car as shown in the example below, has specific values for variables such as the number of cylinders or horsepower. Individual values belonging to an entity are connected by line segments and form a polyline or a Bézier curve [Gruendl 2016].



*Figure 7: Parallel Coordinates with a gradient from blue to red displaying metric scales and categorical scales*

(<https://serhiipuzgrov.com/wp-content/uploads/2019/01/Sk%C3%A6rmbillede-2016-02-01-kl.-18.31.24-768x380.png>)

In expert’s domain Parallel Coordinates are known being capable to display many dimensions and that there is no loss of data due to projections [Blaas 2008]. However, in the information community no more than 8–12 dimensions should be displayed at once.

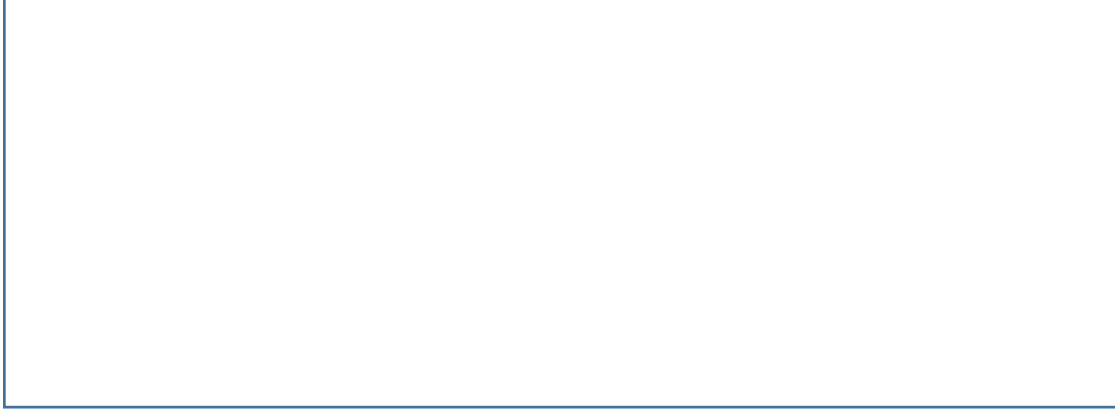
In Fig. 7, the distribution of the red lines can be followed well from the bottom left to the top and down again with the acceleration axis. However, for the last two dimensions, the limitations of the parallel coordinates with respect to categorical data also become noticeable, since they coincide on a few lines that do not reveal a distribution.

The limits of scalability lie in the number of high-dimensional data points and associated lines.

A general comment about Parallel Coordinates compared to other visualization techniques is that the “technique gives a good overview of the data.” [Johansson 2016] This advantage of PCs is related to the strength of PC displays revealing patterns in the distribution as shown Fig. 8, such as different types of

correlations, regions with higher or lower density, textures, and outliers [Johansson 2007, Blaas 2008].

Furthermore, Parallel Coordinates depict different visual patterns for various correlations. The most important correlation patterns are shown below:



*Figure 8: Common patterns in Cartesian coordinates (top) and their dual representation in Parallel Coordinates (below) [Heinrichs 2013, p. 20]*

Since interactivity generally improves the scalability of information visualizations, Edsall (2003) presents in case studies of how exactly inter-activity can improve PCs, while Graham (2003) introduces refinements, such as using smooth curves and spreading out points on axes with discrete values.

Handling of huge data sets, e.g., cohort data is especially given with appropriate interaction techniques [Chan 2006], such as brushing and linking [Heinrich & Weiskopf 2013].

### **Parallel Coordinates for a Broad Audience**

Parallel Coordinates rise in their *popularity* as the number of publications with the term ‘parallel coordinates’ in the title has been rising from 14 in the year 1991 to approximately 543 in 2011, with a total of 5620 publications as reported in 2013 [Heinrich & Weiskopf 2013]. In March 2023 there are 17,200 publications explicit about ‘parallel coordinates’ listed in Google Scholar. Since then, approximately 400 publications have been added to Google Scholar, yielding 17,600 results in May 2023.

As I said earlier, Parallel Coordinates are typically used by data analysis experts in specific domains [Heinrich & Weiskopf 2013], but some authors also mention the potential of this type of visualization being used outside the expert world. Johansson & Forsell (2016) experienced in case studies that “although users tend

to be confused at the beginning they quickly learn how to use Parallel Coordinates and tend to appreciate the way they can interact with their data” [Johansson & Forsell 2016].

It is reasonable to assume that a Parallel Coordinates plot which is “visually appealing with intuitive interactions would attract the attention of more users and stimulate uptake and usage” [Johansson & Forsell 2016]. Therefore, I investigate which design decisions contribute to the aesthetics and interaction as well.

## Challenges of Parallel Coordinates

summarized by Heinrich & Weiskopf (2013):

- Overplotting (too many lines overlapping could occlude important patterns in the data)
- the appropriated order of the axis influences the revealed patterns displayed
- line-tracing gets more difficult with the amount of records
- nominal and ordinal data have to be treated differently than metrical data to avoid confusion
- The *ordering of the axes* is also important. While relationships between adjacent axes can be easily seen, the order of the axis should be well thought.

Addressing these issues with the following ideas and techniques:

- An *introductory example* would help to avoid this confusion and accelerate the understanding of how to interpret Parallel Coordinates.
- To reduce complexity, the number of displayed axes should be *reduced to a very few axes necessary to convey the message*.
- Furthermore, I purpose an *incremental construction* concept to enable the complexity increasing in the tempo of the user.
- Reducing *transparency* is a technique to minimize overplotting. To draw attention to the risk factors in the affected group I assign different transparencies. In the background layer are the healthy participants and further in the foreground appears the affected group.
- In addition, two very *different colors* are used to distinguish between the compared groups. The affected group is shown with a signal color such as red, while healthy patients are assigned a calmer color, e.g., blue or violet.

- the pattern reveals likely to be found in the lower part of the graph, where the values associated with a healthy lifestyle are shown; In addition, incremental construction would reduce the overload of information produced by overplotting. Moreover, I could imagine a fast animated step-by-step plotting of the lines so that the user could see more patterns within the data.
- The *order of the axes* is given by the importance of the risk factors in reading direction.
- *Line tracing* of the protagonist as the focus information is highlighted, and the other data are considered as contextual information for comparison and insight. This concept could be further developed by marking the line picking.

[Heinrich & Weiskopf 2013]

Johansson & Forsell (2016) also propose that it would require *qualitative evaluation* to investigate users' subjective actions and opinions regarding Parallel Coordinates.

In the next section, I will consider time-oriented data displaying with Parallel Coordinates. Aigner et al. (2023) purpose a systematic overview and taxonomy to classify time-oriented data visualization. Here, the underlined attributes show the feature of the SHIP data.

## Time in Cohort Data

The challenge with the time aspect of cohort data was that in this particular data set I used, there were very few time steps over a large time interval of a 15-year period. In addition, the different variables have different numbers of time steps, i.e., only a single time step, some with only two or three time steps with an approx. distance of five years. However, a typical time series represents multiple sample points in time. Therefore, the question is a visualization based on this data can be called time visualization? Moreover, other difficulties I could experience in the user study with this very few time points I reflect in the discussion section.

On this basis, I sought ideas for representing time with Parallel Coordinates on the basis that the general public might be most familiar, i.e., with a simple timeline diagram or life lines visualization.

### 2.4.2 Time-oriented Parallel Coordinates (Temporal PC)

In this section, I will collect different variations of integrating time-oriented data into Parallel Coordinates that appear in the literature. I will discuss advantages and disadvantages of each approach and then examine the potential of each candidate for the general public.

Gruendl et. al (2016) found that there are four basic approaches to consider time in Parallel Coordinates and they also developed a three-dimensional approach.

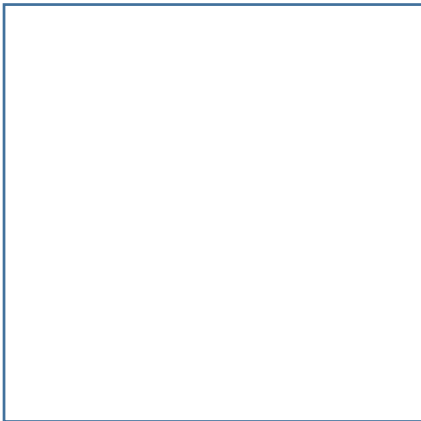
(1) *Time axis*: An additional dimension for time



(1a) The first axis of the Parallel Coordinates shows the time steps. Each line represents a single time step. Lines belonging to one entity, e.g., a person, have the same visual properties, which can be a dotted line vs a solid line [Fig. 9].

Figure 9: Axis *T* depict time steps 0 to 4 [Wong et al. 2016]

The advantage of this variant is that patterns can be made visible. If the values on an axis are sorted in ascending order from bottom to top and all three lines intersect between the two parallel lines, there is a decreasing tendency. If the lines do not intersect, the tendency is increasing, and if all lines accumulate in a single point, it means that the entity has not changed the value of a particular attribute. The limit of the approach to use a single line for each time step means that for  $n$  entities and  $x$  times there are  $n \cdot x$  lines with  $n$  different line properties. That means that this visualization is not suitable for many entities [Wong et al. 2016].



(1b) Another approach with an additional axis is *TimeWheel* [Tominski et al. 2004]. Here, the axes of the dimensions are arranged radially and the time axis is arranged horizontally in the center. The dimensions are limited by the radial arrangement. While in a classical parallel coordinate graphic the lines intersect between two parallel axes, the danger of overplotting is much greater with a TimeWheel visualization [Fig. 10].

Figure 10: TimeWheel [Tominski et al. 2004]

If all lines between a dimension axis and the time axis form a group, then there are additional overlapping sections between these groups of lines. Overall, this visualization would probably overwhelm people who have little or no experience with data visualization. This approach is interesting for a few data entries, but for larger data sets such as that of cohort data, the representation would very quickly lead to serious overplotting.

(2) *Partial time-series:*

In this approach, each dimension has  $x$  axes, one for each time step, plotted separately next to the other time step axes. Here, a data record is encoded by a single line. This would reduce overplotting in comparison to an additional time axis [Grundl 2016]. Secondly, the time series representation is more intuitive due to the chronological order from left to right (Western reading direction). Moreover, time series are already familiar to the general public from the media, such as daily stock price fluctuations [Jiang 2013]. Thus, this variation is a candidate for temporal Parallel Coordinates in the context of narrative visualization.

(3) *One plot per time step:*

Barlow and Stuart (2004) developed the visualization tool *Animator* [Fig. 11] that generates parallel coordinate plots over time. The user can slide through the time slices by a control panel.

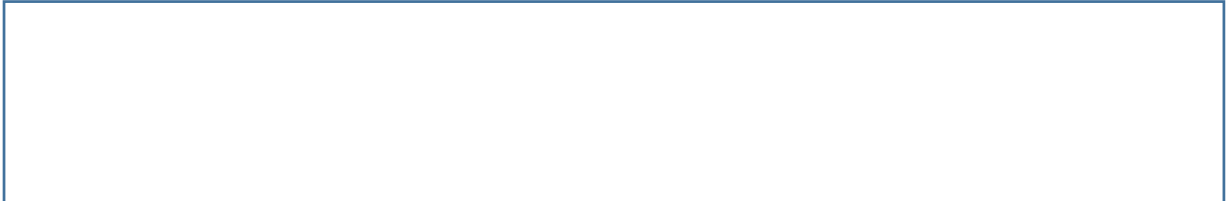


Figure 11: Temporal PC integrated into the Animator dashboard

Theron (2006) had created a similar approach for geo visualization where the user can navigate through a set of stacked parallel-coordinates plots. While the tools can be very supportive for data analyst experts, it still requires a lot of cognitive capacity to compare the data between different time steps. Therefore, it makes the detection of patterns within time difficult and is not suitable for non-experts. The detection of patterns is limited and needs a lot of cognitional afford in resp. of switching the attention between two or more time steps plots and may therefore suitable for experts but not for the general public.



(4) *Coordinated multi-view approach:*

Blaas et al. (2008) propose techniques to enhance the usability of Parallel Coordinates for the exploration of “large, multi-timepoint volumetric data sets”. They name visual clutter as a major problem occurring when visualizing tens of millions of points per time step.

(5) *Three-dimensional Parallel Coordinates:* Blaas et al. employed a focus-and-context visualization where time is integrated as third dimension and displayed as time series in a Parallel Coordinates plot. Between two adjacent axes a time-series plot is integrated with a “pseudo-perspective” with a single vanishing point [Gruendl et. al 2016]. They argue that their three-dimensional approach “emphasize [the] spatial relationship” between dimensions and time, but time is not a spatial dimension and the attributes in a Parallel Coordinates graph often have no spatial reference either. For this reason, it is not very intuitive to display time as a spatial dimension. Furthermore, using three-dimensional plot leads to a high level of complexity that may be appropriate for experts but not for a broad audience.

(6) *Density/Transparency approach,* depict the latest time steps in a high saturated color (as shown below in *Fig. 12* the recent time steps is depicted in signal red) which then is reduces over time [Johansson 2007].



*Figure 12: Transparency approach*

## 2.5 Additional Concepts: Gamification & Cognitive Behavioral Therapy

There is a wide variety of information available on health topics in different media. Most of the information remains very general, such as telling people to lose weight, eat healthier, and exercise more, and the audience remains inactive in absorbing the information from text or video.

I want to actively and emotionally involve the audience in the information by having them actively take choices and do tasks within the flow of the story. I could imagine that with these additional concepts, the tool could support not only self-study, but also clinical treatment.

### Gamification

Since the popularity of computer games rises in all age groups and *learning* is much easier when it is fun, Burgers et al. (2015) note that games are important vehicles for not just learning but also *behavior change*. Gamification is getting more and more involved also in education. Bachen et al. (2016) evaluate character identification for serious game, e.g., to learn how to behave in risky work situations. Identification with a gaming character is a determinant of video game enjoyment [Hefner 2007]. Therefore, it is important to create an appropriated protagonist for the story.

To increase *User Engagement*, I include simple but *engaging interaction tasks* at different points in the story line. This is to ensure that the user's attention continuous to follow the story and to add variety and entertainment to the topic.

Since *Feedback* is a very important concept for both user interface and gaming experience [Burgers 2015], I tried to insert feedback at every point where the user has to take a decision or should perform a task.

There are two types of feedback. *Positive feedback* is a response of the user's correct behavior (i.e., the user chose the correct element). Here, *positive feedback* is implemented in the form of *story progress*, *recognition* from the doctor (character) and *visualization* (for example drawing a line between two selected points in the Parallel Coordinates introductory example puzzle).

In the second scenario, *negative feedback* is given when the user has not clicked correctly or is simply trying out during exploration. Since *motivation* is a huge factor in the psychology of behavioral changes as well as in 'continue playing', I did not make the feedback the opposite to avoid negative emotional responses or decrease motivation. "Negative feedback decreases players' feeling of competence,

but also increases immediate game play. Positive feedback, in contrast, satisfies competence and autonomy needs, thereby boosting intrinsic motivation.” [Burgers 2015]

To avoid that the feeling that he has done something wrong or even feels punished, I have used different *negative feedback techniques*. One solution is to use tooltips to show (humorous) thoughts of the protagonist when a ‘wrong’ link was clicked. Another technique is ‘optional story paths’, such as asking the doctor for more information instead of moving the story forward. Feedback is a form of reinforcement.

From a psychological perspective, Skinner (1971) described the mechanism of positive and negative reinforcement on behavior in his theory of operant conditioning.

### **Operant Conditioning**

The psychology of Operant Conditioning [Skinner 1971] (based on *Classic Conditioning* [Pawlow 1923]) states that a reward *positively reinforces* the behavior that led to it. Skinner (1971) also found in experiments that positive reward reinforces behavior more than negative reinforcement (punishment). Therefore, with a strong identification with the protagonist, the reward of the protagonist’s recovery could increase the user’s *motivation* to change his or her own lifestyle.

### **Cognitive Behavioral Therapy**

Especially later in the story the emphasis is on creating a treatment plan and considering lifestyle or behavior changes. One major risk factor for fatty liver is an inappropriate diet or eating behavior [Yasutake 2014]. *Cognitive Behavioral Therapy* (CBT) is used to treat various emotional and behavioral problems as well as psychological disorders, such as eating disorders [Hazlett-Stevens & Craske 2002]. The general concept of CBT considers both. First, “cognitive processes were [...] recognized as an important domain of psychological distress” [Hazlett-Stevens & Craske 2002]. Secondly, behavior and cognitive processes influence each other. For example, not only can thoughts lead to behavior change, but conversely, behavior change is also important and can alter mental and cognitive processes, e.g., feelings and thoughts. Thus, a combination of cognitive and behavioral therapy was established.

Second, “any given behavior is seen as a function of the specific environmental and internal conditions surrounding it” [Hazlett-Stevens and Craske 2002]. Therefore, the authors suggest that the specific function of a person’s behavior has to be revealed in order to better understand and predict it.

The authors demonstrate the importance and extent of the psychological component on health problems by the example of chronic pain behaviors that “are believed to be maintained in large part by attention from others” [Hazlett-Stevens and Craske 2002].

In psychological treatment practice, a concrete problem or problem area is first addressed, usually by the patient, and reflected upon with the help of the therapist. CBT assumes that cognition, thinking and feeling can be changed through behavioral changes. This means, in general, if patients improve their behavior and acts more appropriate, they will usually feel better in the first place and think more positively and solution-oriented when experiencing personal progress. In practice, there is an emphasis on concrete instructions for action. Concrete suggestions for action are sought in the patient’s individual social environment during talk therapy and are made even more concrete, for example, by means of a written text, in order to consolidate behavior patterns or to recall them in unstable situations (e.g., emergency plans in the case of acute anxiety).

Therefore, the aim of a ‘treatment section’ within a storyline should be to *concretize or illustrate the information* using the patient’s life circumstances as an example. While the user is taking concrete decisions for the protagonist of the story, they are supposed to think about their own lifestyle. This concept would be promising for applications for a general public, as most of these individuals better learn by example and are *not used to think highly abstract*.

Thus, the tool should provide more *application-oriented examples* to ask what the patient can do to achieve improvements. Moreover, while the user puts himself in the patient’s situation and takes lifestyle change decisions, the user learns a strategy and the process how to change behavior.

An important concept in Cognitive Behavioral Therapy is to promote ‘*self-efficacy*’, the inner belief that one can cope well with difficult or challenging situations.

## 3 Data

### 3.1 SHIP Data

For the realization of the data story about ‘Non-alcoholic fatty liver’, we used the SHIP data set – one of the largest epidemiological databases of a population sample. The cross-sectional ‘Study of Health in Pomerania’ (SHIP) collects comprehensive data about participants who live in the Northern and Eastern Pomerania to determine risk factors for diseases [Klemm 2015]. For the data visualization and character design we used variables that have strong correlation with fatty liver. Correlation were taken from previous studies on the SHIP data to identify risk factors for ‘Non-alcoholic fatty liver’.

Epidemiology is concerned with the prevalence, causes, and consequences of health-related conditions and events. Large population studies, such as the Study of Health in Pomerania (SHIP) [Völzke 2011], collect data about participants to screen the risk factors for various diseases. This information is used not only used to “determine risk factors for specific diseases, but also to inform people about healthier lifestyles” [Klemm 2015]. Thus, the intent and message of my prototype is to communicate avoidable risk factors and the importance of a healthier lifestyle to avoid fatty liver.

#### Data Subsets

We considered three data sets in the period between 1997–2012. Individual data sets SHIP-0 (1997–2001), SHIP-1 (2002–2006) and SHIP-2 (2008–2012) are recorded in time steps of five years. The aim is to visualize each variable in these three time steps (first year, 5th year and 10th year).

We would use the following abbreviations for the data subsets:

time step 1: **s0**  $\leftarrow$  SHIP-0 (1997–2001)  
time step 2: **s1**  $\leftarrow$  SHIP-1 (2002–2006)  
time step 3: **s2**  $\leftarrow$  SHIP-2 (2008–2012)

### 3.2 Description of Selected Variables

Two variables are needed to determine the diagnosis of ‘Non-alcoholic fatty liver’. First, we have a variable for the diagnosis made via ultrasound scan of the liver which is recorded in the first and the third time step. “Hepatic steatosis was defined as the presence of a hyperechogenic ultrasound pattern of the liver” [Ittermann 2012]. Second, to exclude participants with an alcohol problem, those with a positive entry in the variable 'alcohol problem' must be removed from the final data set. Risk factors of steatosis hepatis or fatty liver are “sex, age, physical activity, alcohol consumption, waist circumference, and food intake pattern” [Ittermann 2012]. The authors characterize “physical activity, food intake pattern, and alcohol consumption” as lifestyle factors.

Variables required for the story are classified into three main groups:

- (1) Diagnose-related variables
- (2) Demographic variables: For the narrative context, we need variables that can be used to describe the *personality and personal background of protagonist*, e.g., gender, age, family, education, occupation,
- (3) Risk factors or lifestyle-related variables, e.g., *waist-to-height ratio, body mass index, physical activity, food intake pattern, and alcohol consumption*
- (4) Liver function values to describe the health status and degree of the disease, e.g., inflammation, damage of the liver (increasing liver enzymes)

In addition, for each variable of group (2) and (3) I assign a *normal range* or *reference range* which is visualized by differently colored axis intercepts. The information is taken from the “Laboratory catalog of Greifswald University Medical Center of the Ernst Moritz Arndt University – KdÖR -, Institute for Clinical Chemistry and Laboratory Medicine”.

#### Protagonist-related Data and Risk Factors

I derived the *waist-height ratio (whtr)* from waist circumference divided by height as this value is more independent of the person’s muscle mass and a “better screening tool than waist circumference or BMI for adult [...] metabolic risk factors” [Ashwell et al. 2012]. However, to test which term (BMI or waist-height ratio) is more familiar or understandable to the target group, I tested both variables as risk factors.

A Waist-to-Height Ratio equal or greater than 0.5 “identifies more people at ‘early health risk’” [Ashwell et al. 2012]. Herpertz (2009) associates a normal weight with a ratio between 0.4 and 0.45. Just as underweight is not considered beneficial to health, there must also exist a lower boundary for the Waist-Height Ratio. The “optimal distribution of nutritional status” is found in the group with a Waist-Height Ratio between 0.4 and 0.5. where the highest proportion of normal weight (BMI) was found. Therefore, WHtR is an indicator of nutritional status and early health risk” [Zhang et al. 2016]. Thus, there is a normal or health-promoting range between 0.4 and 0.5.

*Waist-to-Height Ratio* is calculated as:

$$whtr = \frac{waist}{height} = \frac{som\_tail\_x (cm)}{som\_gro\_s2 (cm)}, \text{ where } x = s0, s1, s2 \text{ for time steps 1, 2, 3}$$

Table 1: WHtR – Nauheimer-Adipositas-Skala [Herpertz 2009]

WHtR group	Waist circumference/height	SHIP data
Adipositas Class III–V	$\geq 0.73$	max: 0.78
Adipositas Class II	0.64–0.72	
Adipositas Class I	0.55–0.63	
Overweight	0.46–0.54	
Normal	0.4–0.45	
Underweight	$< 0.4$	min: 0.33

According to A. Quetelet *Body Mass Index* is calculated as follows

$$bmi = \frac{body\ weight\ (kg)}{body\ height^2\ (m^2)} \quad [\text{Schorb 2022}]$$

A person is considered to be of *normal weight* if her/his BMI has a value between 18.5 and 24.9 [Oliveros et al. 2014]. To subdivide the axis more precisely, I could use the following classification. The highest ratio with 0.78 lies within the Adipositas Class III. For simplicity, I combine the upper extreme Adipositas groups into classes III–V.

Table 2: Body Mass Index Classification [Oliveros 2014]

BMI group	(kg/m <sup>2</sup> )	Bins	SHIP data
Adipositas Class III–V	$\geq 40$	6.5	max: 48
Adipositas Class II	35.0–39.9	4.9	
Adipositas Class I	30.0–34.9	4.9	
Overweight	25.0–29.9	4.9	
Normal weight	18.5–24.9	6.4	
Underweight	$< 18.5$	2.4	min: 17

Table 3: Fatty Liver Groups

stea_s0	stea_s2			Participants that 1 = do not recover from fatty liver 2 = get a fatty liver 3 = recover from fatty liver 4 = stay with a healthy liver
		1	0	
	1	1 --	3 - → +	
	0	2 + → -	4 + +	

To convey the **risk factors**, group 1 and 2 are interesting. However, for the presentation of lifestyle change, only group 3 is important, since all individuals within this group **recovered** from fatty liver.

## Lifestyle Change

The story’s message should be that a healthy lifestyle or lifestyle change avoids risk factors for fatty liver. ‘Lifestyle change’ means that a person learns to adapt her or his behavior “to move toward a state of optimal health” [O’Donnell 1986].

Ittermann (2012) identifies “physical activity, food intake pattern, and alcohol consumption” as lifestyle factors associated with steatosis hepatitis in the SHIP data set.

However, an identification of appropriated variables is not clearly delineated. For example, a reduction in BMI and waist circumference may have been achieved by a healthier diet combined with exercise, but may also have been caused by a disease or substance abuse.



A combination of variables could be helpful in determining whether a subject is behaving healthily, but uncertainty remains due to individual life circumstances that are not represented in the data.

Thus, I decided to choose a combination of variables: *Weight loss*, *physical activity*, *lower alcohol consumption* and *quit smoking* and exclude subjects who smoke or started smoking for example.

Physical activity besides eating behavior yields as major lifestyle factor related to obesity and diabetes mellitus type 2 [Wing et al. 2001]. The variable is available for the first and the third time step. Thus, a change in lifestyle is particularly plausible when a person does not exercise in the first time step but in the third step (physact\_s0 = 0 AND physact\_s2 = 1).

Diet would be an important variable in this context but unfortunately for this purpose the SHIP data set only contains data from the first time step. As Wing et al. mentioned, not only adaption is important but also maintenance of the lifestyle change. Therefore, I used ‘smoking status’ with value ‘1 = Ex-Smoker’ to find subjects that already have changed their lifestyle.

Physical activity (during leisure time) is defined as:

(0) = no or little sport

(1) = regular or much sport

“How often do you do sports in the summer?” (SPORTSO1) and

“How often do you do sports in winter?” (SPORTWI1)

1 = Regularly more than 2 hours per week

2 = Regularly 1 to 2 hours per week

3 = Less than 1 hour per week

4 = No sport activity

if (SPORTWI1 != (1 or 2) **AND** SPORTSO1 != (1 or 2))

then: code = 0 (no or little sport)

if (SPORTWI1 = (1 or 2) **OR** SPORTSO1 = (1 or 2))

then: code = 1 (regular or much sport)

This variable divides physical activity very roughly into two groups. Accordingly, people with very different levels of activity fall into one group, e.g., a person who exercises more than 2 hours per week in summer and winter is treated in the same way as someone who only exercises one hour per week in summer. I would suggest more categories for more precise specification. The definition has to be included in the story as additional information.

Nevertheless, physical activity is an interesting variable to see lifestyle change, as most people that return their fatty liver to a healthy liver in the last ultra-sonic examination started exercising (41.13 %) or keep exercising (28.48 %), in summary 69.61 %.

### Diagnose-related Variables

Besides the variables for the diagnosis via ultrasonic imagery (stea\_s0, stea\_s2) liver function values from blood test provide additional information about the actual health status of the liver. I looked up the reference range for each variable in the “Laboratory catalog of Greifswald University Medical Center of the Ernst Moritz Arndt University – KdÖR –, Institute for Clinical Chemistry and Laboratory Medicine” where critical values are highlighted.

Table 5: Liver Function Values from SHIP-0 to SHIP-2  
(see Appendix for detailed Table 5.2)

Variable (natural language)	Label	Datatype	Value of the Subject (Protagonist)	Reference Range
<b>Liver Function Values</b>				
Triglyceride (TG)	tg_s_s0	numerical	2.63	< 1.9 mmol/l
	tg_s_s1		2.67	
	tg_s_s2		2.57	
Gamma-Glutamyl-Transferase (GGT)	ggt_s_s0	numerical	0.44	< 0.96 $\mu$ katal/l
	ggt_s_s1		0.52	
	ggt_s_s2		0.93	
Aspartate aminotransferase (ASAT) Alternative name: glutamic oxaloacetic transaminase (GOT)	asat_s_s0	numerical	0.27	< 0.59 $\mu$ katal/l
	asat_s_s1		0.21	
	asat_s_s2		0.16	

\* Values were taken from: Laboratory catalog of „Universitätsmedizin Greifswald der Ernst-Moritz-Arndt-Universität – KdÖR –, Institut für Klinische Chemie und Laboratoriumsmedizin“  
(<https://www2.medizin.uni-greifswald.de/klinchem/aktuell/leistungsspektrum/Laborkatalog>)

To reduce cognitive overload, I selected three values from the table for the story because these variables are most frequently cited as important liver function values and are included in all three time steps: *Triglycerides*, *Gamma-glutamyl transferase* (GGT), and *Aspartate aminotransferase* (ASAT).

### 3.3 Preprocessing, Handling of Missing Values & Outliers

As we want to consider only ‘Non-alcoholic fatty liver’ I selected subjects who have *no alcohol problem*. Secondly, for the time-oriented visualization we need consistently available data in the *time steps*. Thirdly, all subjects who have *missing values* in the variable for the diagnosis of fatty liver, the selected variables for the risk factors (variable group 2) and liver function values (variable group 3) were deleted. Furthermore, we removed *extreme outliers* that would effectively compress the rest of the distribution to a very small area. That would make it difficult especially for non-data specialists to read the data. Another problem here would be the subjective perception of values compared to others. When I recognize an extreme outlier, other values that are also outside the normal range seem much less dramatic.

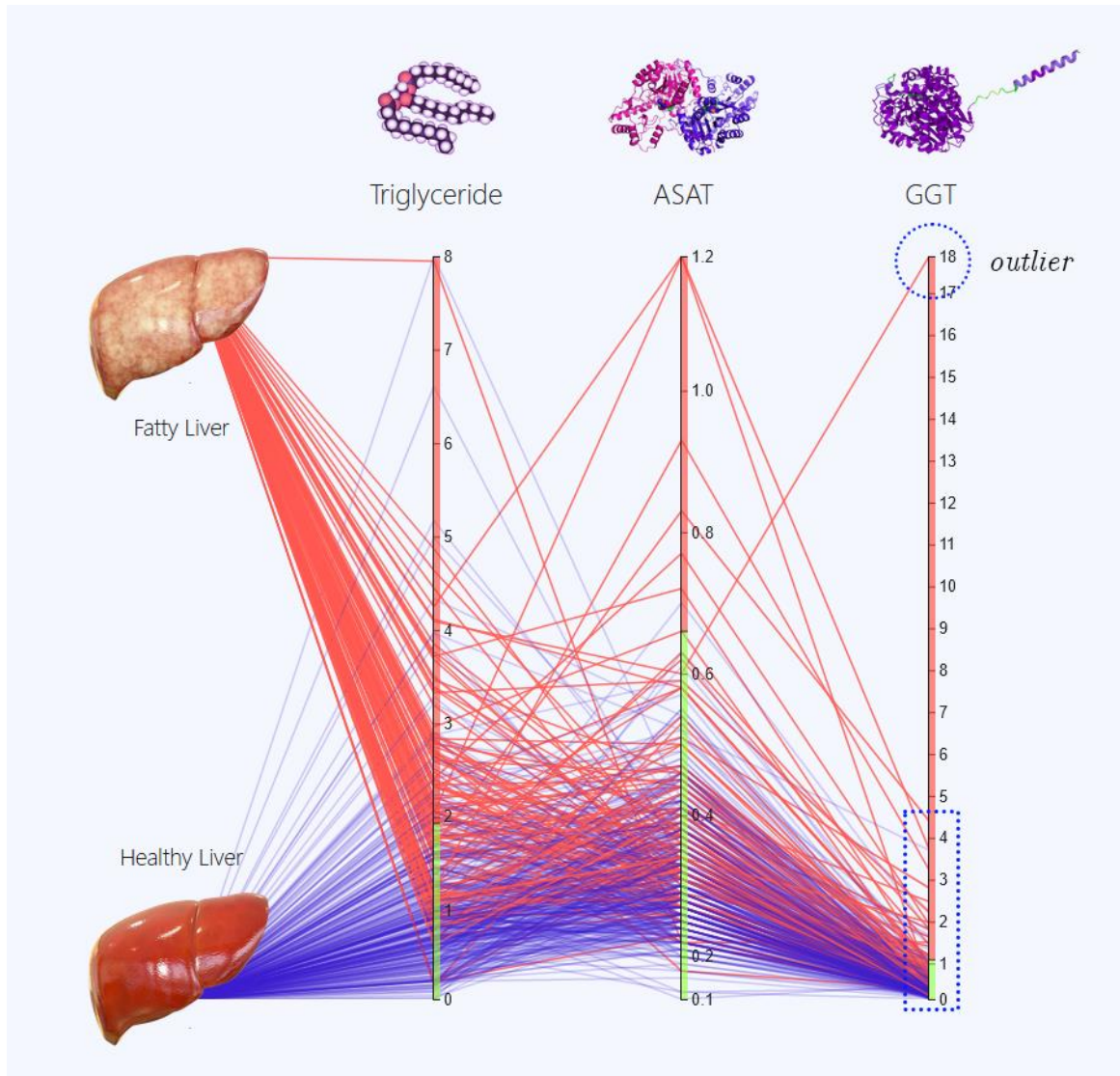


Figure 13: Example: Outliers influences the distribution of the axis and recognition of the values [early example in the design process of the PC plot]

Table 5: Liver Function Values from SHIP-0 to SHIP-2

	Documentation of cleaning the data	Removed	Total
	Original data set	–	<b>4309</b>
(I)	Non-alcoholics: After the first step, the original table was reduced from initial 4309 to 3114 subjects as 1172 have been declared having an alcohol problem and 22 did not report.	1194	<b>3114</b>
(II)	Participants in all three time steps: Then 712 subjects did not participate in the second time step and from persons that have taken part in the first two time steps 1100 do not come to the third examinations. Thus 1302 subjects from the non-alcoholics take part in all three times steps and are therefore candidates for the time-oriented visualization.	1812	<b>1302</b>
(III)	Removing subjects with missing values: But I still have to filter the data due to missing values in (1) diagnosis of steatosis hepatis (fatty liver) to 1278, and in (2) risk factors: bmi [-4], whtr [-3], ffs [-5], diabetes [-0], alcohol consumption [-94], smoking [-0], as well as in the (3) liver function values: I detect 640 missing values in the second time step of variable GGT_s1 and very similar in the variable for ASAT_s1. (3.1) contains a table with 1136 subjects that have no missing values in the first and third time step. In the version (3.2) I delete all missing values in the mentioned variables	(1) 24 (2) 106 (3.1) 36 (3.2) 640	<b>1278</b> <b>1172</b> <b>1136</b> <b>496</b>
(IV)	Outliers: The <b>risk factors</b> table (2) has one extreme outlier only in alcohol consumption ( <i>g/day</i> ). In the first time step there is a man with 630.45 <i>g/day</i> where the next data point is 212.4 <i>g/day</i> . For the <b>liver function value table</b> “Version Two time steps” (3.1) “Version Three time steps”: For the version with considering only two time steps, namely the first and the third, I have created a separate table. In tg_s0 two subjects with 19.66 <i>mmol/l</i> and 16.87 <i>mmol/l</i> were removed due to the gap to the next value at 9.58, then the same persons with 25.64 $\mu\text{katal/l}$ and 17.88 $\mu\text{katal/l}$ GGT value were deleted and in ggt_s1 a subject with 19 $\mu\text{katal/l}$ . So here we have to delete five rows as well. (3.2) I removed five subjects due to extreme outliers. Namely, there was high triglyceride values in tg_s0 16.87 <i>mmol/l</i> where the next subject has 7.93. <i>mmol/l</i> , as well as a man with 14.27 <i>mmol/l</i> in time step 3. With ggt_s0 there are two persons having 25.64 $\mu\text{katal/l}$ and 17.88 $\mu\text{katal/l}$ where the next value is 6.7 $\mu\text{katal/l}$ . and in ggt_s1 a subject with 19 $\mu\text{katal/l}$ was removed. 2.61 $\mu\text{katal/l}$ in asat_s0 was deleted.	(2) 1 (3.1) 6 (3.2) 6	<b>1171</b> <b>1130</b> <b>490</b>

We decided to take the data set with subjects who participate in all three time steps, that means the original data set of 4309 subjects was reduce to 496 in respect to removed outliers there were 490 subjects left. Of these 496 subjects, 61.69 % were women, and the age range of participants is 20 to 78 years in the starting sample.

### Preprocessing for d3.js

We use the JavaScript library d3.js (Data Driven Documents) to support our data visualizations. For the processing with the d3 library, the data has to be converted into a specific csv format that is required for the d3 parser. To do this, all decimal separators must be decimal points. All textual data values such as ‘US pos.’ have to be replaced by numerical values. Finally, all variables and data values have to be separated by comma.

For example, diagnosis with three liver function values:

<i>Variables/Dimensions</i>	stea_s0,tg_s_s0,gggt_s_s0,asat_s_s0
<i>Data record/Values</i>	1,0.51,0.2,0.62

The final csv file can be uploaded to a GitHub repository which has to be ‘public’ to be accessible for d3 library. A link to the raw format of the table has to be included into the JavaScript file.

### 3.4. How to find a Protagonist in Data?

Since the story should be data-driven and credible, I based the development of the storytelling and character design on the data of a selected subject.

A very early lesson when analysing the data set was that there is no ‘perfect protagonist’. Therefore, I decided to define a query filtering schema to get as close as possible to this “perfect candidate”. I divide the query in two steps:

First, conditions that have to be fulfilled (*“must have”*), *i.e. Hard criteria*. Second, I define some soft criteria to choose the most ideal subject among those (*“nice to have”*), *i.e. Soft criteria*.

*Hard criteria* are connected by conjunction (AND). For a protagonist candidate by the example of Non Alcoholic Fatty Liver all criteria have to be logically true:

First, the protagonist has a fatty liver at the beginning and no fatty liver at the last time step (positive diagnosis in time step 1 and a negative diagnosis in time step 2).

Second, the story is about Non Alcoholic Fatty Liver (NAFLD), so all persons with an alcohol problem are excluded (exclude unfavorable variations of the disease).

Third, since too much visceral fat is the main risk factor for fatty liver, the person should be overweight in the first time step and then lose weight in the following time steps (include candidates with an improved change in the main risk factor).

Fourth, the candidate has changed her/his lifestyle and behavior, e.g. started exercising and do not smoke in the last time step or is an ex-smoker or even quit smoking (include candidates with a lifestyle change).

Thus, for the minimum conditions yield:

CANDIDATES\_ *hardcriteria* = subjects S where

(*stea\_s0* = US pos. AND *stea\_s2* = US neg.) AND  
 (OVERWEIGHT = True) AND (WEIGHT-LOSS = True) AND  
 (*alcproblem\_s0* = 0) AND (*physact\_s2* = 1) AND (*smoking\_s2* = 0|1)

OVERWEIGHT: if (*whtr\_s0* > 0.5) then True

WEIGHT-LOSS: if (*som\_tail\_s0* > *som\_tail\_s2*) then True

Legend 1 (Hard criteria):	
<i>stea_s0, s2</i>	Steatosis hepatis = Fatty liver
<i>whtr_s0</i>	Waist-to-Height Ratio
<i>som_tail_s0</i>	Waist circumference
<i>alcprobl_s0</i>	Alcohol problem
<i>physact_s2</i>	Physical activity
<i>smoking_s0</i>	Smoking status

After filtering the data set, 17 subjects remained out of 4308.

The *soft criteria* then use a disjunction (OR) to find an optimal subject among the candidates.

At the beginning I tried to find a *male* protagonist, because fatty liver is more common in men than in women. I could imagine creating a prototype for men with a male protagonist and one for women with a female protagonist, because the identification would then be higher.

The problem to find a male protagonist was: First, the hard criteria do not fit, only three men come near to a normal weight in the last time step. Secondly, two of them remain smokers and we do not want to propagate smoking. The one man that does not smoke has a very little difference between start weight and end weight which does not tell the story of losing weight as life style change. For this reason, I chose a female protagonist who has an overall constant weight loss.

In addition, women have an indirect influence on the lifestyle of men. In the role of a partner or wife, she can influence for example the physical activity of her partner or what and how much is bought, cooked and finally eaten. In addition, a partner can also change the other's attitude through conversation. So, if a wife sees the information that men are very much at risk for fatty liver, she would try to change her husband's lifestyle out of love and care. Therefore, this medium should help to empower women.

So, more specifically, the weight loss should bring the person to a healthy weight in the range of *normal weight* in the last time step, i.e., a BMI between 20–25 and a *waist-to-height ratio* below 0,5. These results would make the story much more interesting because the proband respectively the protagonist has achieved an optimal healthy weight at the end.

There is a variable for the balance of diet *Food Frequency Score* (*ffs\_s0*) that is only available for time step 1 (See description in: Dokumentation für die Definition und Nutzung transformierter Variablen in SHIP-0, 2012). Moreover, explaining the composition of the variable especially to broader audience could easily lead to information overload. Nevertheless, in respect to the storyline I could imagine that the proband has an unbalanced/poor nutrition at the beginning. But at the end you cannot say anything about the state of nutrition that is based on data or you have to add a fictions part, e.g. that the dietary is more balanced and healthy.

As diabetes mellitus type 2 is a risk factor for fatty liver, it would be great to show that even a diabetic can cure his fatty liver. However, there is not a single person in SHIP-data who recovered from diabetes. The fact that a person started exercising is a strong twist in the story, but perhaps it is more realistic for a person to already do some sports and then increase training, as I have seen in the series “The Nutrition Docs” (protagonist: Carlo Bourdieu).

Furthermore, as alcohol consumption is still a risk factor even when the person has no alcohol problem, I prefer a subject that reduces drinking alcohol.

$\{ \text{CANDIDATES\_hardcriteria} \} \rightarrow \{ \text{CANDIDATES\_softcriteria} \}$

where: Softcriteria are

(*whtr\_s0* > 0.5 AND *whtr\_s2* < 0,5) OR  
 (*som\_bmi\_s0* > 25 AND *som\_bmi\_s2* < 25) OR  
 (*som\_bmi\_s0* > *som\_bmi\_s2*) OR  
 (*som\_tail\_s2* > *som\_tail\_s1* > *som\_tail\_s0*) OR  
 (*physact\_s2* = 1) AND (*smoking\_s2* = 0) OR  
 (*alkligt\_s0* > *alkligt\_s2*)

Legend 2 (Soft criteria):	
<i>female_s0</i>	gender, sex
<i>som_bmi_s0, s2</i>	Body Mass Index (BMI)
<i>som_tail_s0, s2</i>	Waist circumference
<i>whtr_s0</i>	Waist-to-Height Ratio
<i>physact_s2</i>	Physical activity
<i>diabetes_s0</i>	Diabetes
<i>alkligt_s0</i>	Alcohol problem
<i>smoking_s0</i>	Smoking status

For my final candidate, I selected a subject who meets most of the conditions. A woman who meets two most important conditions for successful weight loss, namely: First, a reduced Body Mass Index, where *som\_bmi\_s0* = 27.3 > 25 AND *som\_bmi\_s2* = 23.3 < 25, which is in normal weight range.

Secondly, it is very interesting that her BMI in time step 2 is already in the normal weight range at 24.4, but her waist-to-height ratio is not. It decreases from 0.55 to 0.51 and finally in time step 3 it is below 0.5 with 0.49.

Not only does she steadily decrease her waist circumference over time (88.4 cm > 82.5 cm > 78.5 cm), but she also reduces her alcohol consumption from 3.36 to 2.28 and abstinence in the last time step. It is also beneficial that she is an ex-smoker as this shows that she was able to change her lifestyle to a healthier one.

Furthermore, *liver function values* are interesting to see a process in getting a healthy liver. Triglyceride, ASAT and GGT, LDL Cholesterol and HDL Cholesterol are liver function values. We can look for values in different time steps and if they are out of the reference range in the first time step and getting closer to the reference range or being in the reference range at the end.



Here, are two example of the soft criteria for liver function value, but other variation is also possible. s0 stands for the first time step, and s2 for the last time step.

Variation 1:

(variable\_s0 > reference range AND variable\_s2 < reference range)

*example:* (tg\_s\_s0 > 1.9 mmol/l AND tg\_s\_s2 < 1.9 mmol/l)

Variation 2:

(variable\_s0 > variable\_s2 > reference range )

*example:* (tg\_s\_s0 > tg\_s\_s2 > 1.9 mmol/l )

Variation 3:

(variable\_s2 > variable\_s1 > variable\_s0)

*example:* (tg\_s\_s0 > tg\_s\_s1 > tg\_s\_s2)

Legend 3 (Soft criteria):	
<i>female_s0</i>	gender, sex
<i>whtr_s0, s1, s2</i>	Waist-to-Height Ratio
<i>som_bmi_s0, s1, s2</i>	BMI (Body mass index)
<i>physact_s0</i>	Physical activity
<i>tg_s_s0, s2</i>	Triglyceride
<i>ggt_s_s0, s2</i>	Gamma-Glutamyltransferase
<i>asat_s_s0, s2</i>	Aspartat-Aminotransferase

Based on this filtering comprising hard and soft criteria, I found a representative protagonist (See Table 4 and Table 5).

Table 4: Selection of Variables from SHIP-0 to SHIP-2  
(see Appendix for detailed Table 4.2)

Variable (natural language)	Label	Datatype	Values of the Protagonist (no 2013160005002)	Description and Measuring unit
General				
Gender/sex	<i>female_s0</i>	boolean	1	0 = male, 1 = female
Age	<i>age_ship_s0</i>	ordinal	65	<i>years</i>
	<i>age_ship_s1</i>		70	
	<i>age_ship_s2</i>		77	
Year of Birth ( <i>derived variable</i> )	<i>year_birth</i>	ordinal	1933	Derived by [year ← exdate_ship_s0] – [age_ship_s0]
Family				
Marital status	<i>marit_s0</i>	categorical	2 = married, living separately from partner	↓
1 = married, live with partner, 2 = married, living separately from partner, 3 = single, never been married, 4 = divorced, 5 = widowed, 8 = do not know, 9 = refusal to answer				
Birth	<i>birth_w_s0</i>	boolean	1 = One or more children	at least one birth 0 = No, 1 = Yes
Education				
Educational level	<i>edlevel_s0</i>	categorical	6 = Abitur	↓
1 = Hauptschule, 2 = Volksschule, 3 = derzeit Schüler/kein Abschluss, 4 = Mittlere Reife, 5 = Realschule, 6 = Abitur, 7 = Fachabitur, 8 = Fachhochschulreife				
Occupation				
Occupational group	<i>hf_kldb_75_s0</i>	categorical	864 = kindergarten teacher, childcare worker	KldB 1975 (Classification of occupations)
Somatic Characteristics				
Height	<i>som_groe_s2</i>	numerical	160	<i>cm</i>
Weight (Body weight for s0 and s1 are derived from bmi_s0 and bmi_s1)	<i>derived for s0</i>	numerical	69.8	<i>kg</i>
	<i>derived for s1</i>	numerical	62.4	
	<i>som_gew_s2</i>	numerical	60.6 (ideal weight)	
Diagnosis				
Fatty Liver ( <i>lat. Steatosis hepatis</i> )	<i>stea_s0</i>	boolean	US pos.	US pos. = Fatty Liver, US neg. = No Fatty Liver, Diagnosed with ultrasound imaging
	<i>stea_s2</i>		US neg.	

Risk Factors for Fatty Liver				
BMI – Body Mass Index	<i>som_bmi_s0</i>	numerical	27.3 (overweight)	See Table 2: Body Mass Index Classification
	<i>som_bmi_s1</i>		24.4 (normal weight)	
	<i>som_bmi_s2</i>		23.3 (normal weight)	
Waist circumference	<i>som_tail_s0</i>	numerical	88.4	<i>cm</i>
	<i>som_tail_s1</i>		82.5	
	<i>som_tail_s2</i>		78.5	
Waist-to-Height Ratio (derived variable)	<i>whtr_s0</i>	numerical	0.55 (overweight)	$whtr = \frac{som\_tail\_s0}{som\_groe\_s2}$ 0.4 < healthy whtr < 0,5
	<i>whtr_s1</i>		0.51 (overweight)	
	<i>whtr_s2</i>		0.49 (normal weight)	
Alcohol Problem	<i>alcprobl_s0</i>	boolean	0	0 = No alcohol problem 1 = Alcohol problem
Alcohol Consumption	<i>alkligt_s0</i>	numerical	3.36	Alcohol intake <i>g/day</i> (last week) [Alte 2004], if <i>alkligt</i> = 0 then abstinent
	<i>alkligt_s1</i>	numerical	2.28	
	<i>alkligt_s2</i>	numerical	0 = abstinent	
Smoking	<i>smoking_s0</i>	categorical	1	0 = Never Smoker 1 = Ex-Smoker 2 = Current Smoker
	<i>smoking_s1</i>			
	<i>smoking_s2</i>			
	<i>onsetsmok_s0</i>	ordinal	17	years, Age of smoking onset (first smoking)



## 4 Design Concept

With this work, I suggest a design concept to create a cohort data story for a broad audience. The design concept brings together the storyline, character design, visual design of the data visualization and interaction design (User Experience Design). I used a combination of three genres based on Segel & Heer (2010) ‘Interactive Slideshow’, ‘Annotated Chart’ and partly ‘Video/Animation sequences’.

### 4.1 Target Group Characteristics

For the issue of fatty liver, it can be deduced that the age groups 30-89 are affected by fatty liver disease. During the period of 2008–2012, in the age group between 30–49 the risk of fatty liver is one-third, and in the 50–69 old age group, it is almost a half. 70–79 old subjects also have a high prevalence with 38 percent. However, there is a high increase in 30–69 year olds, so that would be the main target group by age.

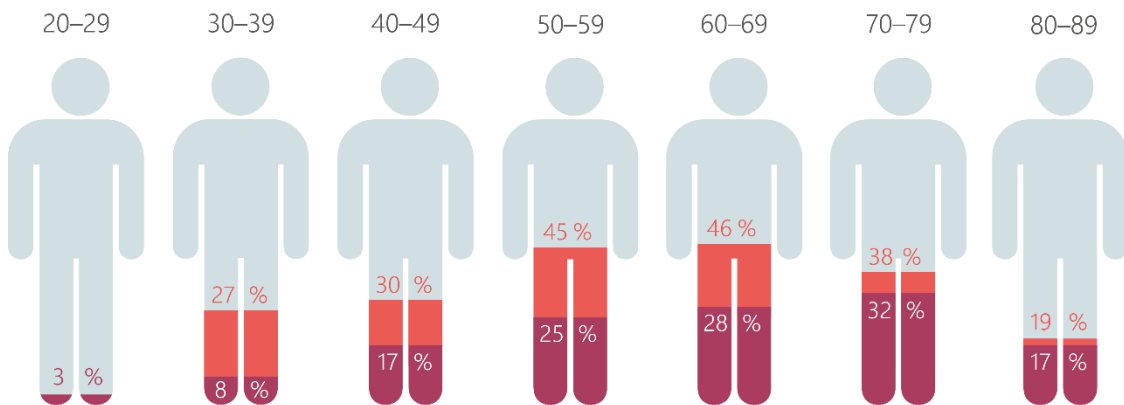


Figure 15: Isotype Chart created from SHIP data: “Increase in fatty liver disease within age groups (1997–2012)”, [work from my prototype]

Other criteria for the target group are middle to higher educational level that is associated with a generally curious and open to learn new things. We have to assume that they are able to work physically with a computer and have at least a little interaction experience with user interfaces or websites.

Expectations of an individual from the general public may clarify regarding the information and layout arrangements, comprehensiveness and support, accuracy to a certain degree, reduced and simplified information, a higher proportion of illustration (images, info graphics etc.) than textual information.

## 4.2 Design of Protagonists

### *Main Character – The Patient*

The protagonist is a person that is affected by a certain health issue the story is about, here fatty liver. These characters should transport the content *emotionally* to the viewer.

A pretest showed how important the protagonist is to users. My first version began with the protagonist briefly introducing herself. This was followed by a digression about the SHIP dataset. Users have reported feeling “abandoned” because the protagonist suddenly disappeared. Therefore, a better idea is to let the protagonist speak about the study.

For the realization of the protagonist’s behavior and dialogs based on the NDR documentary series “Die Ernährungsdocs\*<sup>1</sup>”. They made several reports about fatty liver and other diseases related to avoidable risk factors and lifestyle change. (\*<https://www.ardmediathek.de/sendung/die-ernaehrungs-docs/Y3JpZDovL25kci5kZS8xNTMw>)

I could observe different emotional reactions from patients who get diagnosed with fatty liver for example. In this case, first *fear* and then *sadness* were the main emotions, regarding further liver diseases up to liver cancer and even death. Sadness or depressing could occur when people are confronted with the self-responsibility of their health status. Some people acknowledge this self-responsibility and turn it into a *happy* moment when they agree to change their lifestyle.

In addition, patients sometimes express *disgust* when they look at representation of a fatty liver, or they feel *angry* when confronted with the treatment plan, e.g., staying away from favorite foods or increasing their physical activity.

The main character appears at the beginning, in the middle and at the end of the story in animated video sequences. At the beginning, the character introduces her/himself by talking to the viewer – telling its name, age and diagnosis and feelings about this diagnosis.

### ***Side Character – The Doctor***

A second protagonist is a physician who diagnoses the main character with a fatty liver and examines his health (blood and liver function values) as well as giving advices for treatment. Here the character and role of the doctor has to fulfil certain expectations of the users. Therefore, I visited *jameda.de*, a website where physicians present themselves and are rated by patients.

I collected positive attributes in patient’s ratings. The most common attributes are “competent” and “friendly”. Therefore, I ordered different attributes into these two groups.

(1) Competence-related attributes: “competent”, “knowledgeable”, “very professional”, “well-informing” and give “safeness”, “looks very closely”, “great help”, “comprehensive treatment”, “expertise”, “details”

(2) Social-skills-related attributes also play an important role in forming a personal relation: “friendly”, “personable”, “honest”, “top consultation” → communicative, “approachable”, “takes time”, “listens”, “dedicated”, “motivated”, gives the feeling of being “in good hands”, gives “clear and well understandable answers” → *empathetic* and *understanding*

(3) Not so often mentioned are characteristics that express the doctor’s ability to address and motivate the patient. The expression “honest” is a sign that the doctor talks to patients about how serious their health situation is.

Thus, the role of the doctor in the story is not only to be competent and friendly, but also to *motivate* the patient and get them to change their lifestyle. This starts with being honest about how serious the patient’s health situation is or can become, and then providing motivation by showing a path of healing (treatment plan) while *appealing* to the patient’s self-responsibility.

### ***Just an Idea: Mascot – The Liver***

The main subject of the topic is the liver. As mentioned, a problem with fatty liver is that the liver normally does not cause pain and is not that present in human’s awareness as other health problems. Therefore, the risk of fatty liver, which limits liver functions and can ultimately lead to death, is easily underestimated. The question is how the liver can be more present as one of the most important organs in the human body. To make the liver more present, approachable, even sympathetic, the liver could get a narrative role in the story. One consideration to realize this is to abstract the liver to a mascot with a face that expresses different emotions. The main protagonist can *talk to* the mascot liver and gets *feedback* when something changes or liver health improves.

### 4.3 AI-Assisted Character Design

My aim is to create a character for the patient that the user can identify or empathize with. Analogous to the concept of telling a story with *real data*, I came up with the idea of portraying a protagonist as *vidid* as possible. So, the representation should be close to a real person (or semirealistic as a gaming character for example), but also inexpensive and easy to create for, e.g., data analyst without special design skills.

There are various ways to portray a protagonist. The simplest way to represent a person's character is to write a *textual description* of its appearance, personality, personal background and behavior. We will later come back to this point. First, we want to examine different ways to portray a protagonist *visually*.

While a *photo* of the protagonist is a realistic and an inexpensive representation, it is also a static perspective and cannot bring the character to life, as it is the case with documentaries and reports in which a person's story is presented on *film* [e.g., The Nutrition Docs]. This cinematically representation is very realistic and individual, but also very expensive and not suitable for a more automated generation of data stories.

#### Animated Speaking Character

So I thought of a representation that had the potential to combine the advantages of both approaches – photo and film. Based on these considerations, I searched for a method of a character realization that is somewhere between static imagery and cinematic representation, an “animated photo”. I researched in the field of computer animation for “*animated speaking person/character*” and found several AI tools that could be used to animate a photo or a computer-generated image. Moreover, with state-of-the-art AI tools, it is possible to create an animated speaking character with any given input text.

I decided to use a semirealistic image of the character, because it highlights both the fact that the story is created by real data but the character is still fictive.

With this concept I can combine the advantages of the two approaches described above:

- (1) the figure is brought to life (by the animation),
- (2) there is the opportunity to combine fictitious photorealistic images of patients with real photographs doctors for example without stylistic inconsistency,
- (3) the design of the fictitious character may emphasize their personality,
- (4) the implementation is less expensive compared to a film and



(5) can be realized without special design skills

My original inspiration to design the story characters comes from game design where characters appear very unique and important features of their personality are emphasized. Although these characters are not living human beings, like real persons, a very strong identification with the computer-generated character is possible [Hefner 2007].

### ***Create a Character***

#### ***a) Describe the Features of the Character from Data***

As I described in section 3.3 Variables for the Main Protagonist (Patient Role) I chose a woman with the following attributes derived from data:

A *65-year-old woman* (in 1997 when the data was collected) born in *1933* – the year the Nazis takeover power. She was 6 years old when the Second World War in 1939 started and 12 years old when the post-war period with the “years of famine” began.

She went to school for 13 years and graduated from *high school*, so you can say that she is *well educated*. After school she worked as a *kindergarten teacher* and became a *mother*. Therefore, it is plausible that she *loves children* and has a *social and friendly personality*. She is also married, but lives separately from her partner.

Her height is *160 cm*. At the beginning of the study, she is diagnosed with *fatty liver* and is slightly *overweight* with a BMI of over *27* and an abdominal circumference of *88.4 cm* (measuring point: 06/25/1998) that is an unhealthy Waist-to-Height Ratio of 0.55. A Waist-to-Height Ratio of about 0.5 is associated with certain other disease as cardio-vascular diseases.

She had started *smoking* at the age *17* and had quit smoking and is abstinent to this today (*ex-smoker*). She was able to change unhealthy behaviors (her lifestyle) in the past. So, it can be said that she has a *strong will, a strong personality*.

She does exercise, but probably not enough, as she is overweight at the beginning of the measurement. The *physical activity* variable has only two values and the intensity is not shown here, so it may be that she meets the minimum requirements, which is to exercise once a week and only in either summer or winter. In addition, diet is also very important for the amount of visceral fat. Her food score value describing her dietary versatility is 21, meaning her diet is *in the lower range of a good dietary* pattern defined by a food value between 19 and 30.

For the storyline, we can assume that she increased her physical activity in combination with diet improvement to achieve a healthy weight. As a

consequence, she loses weight in time step 2 and 3 and appears matured, but still appears very much alive.

At beginning of the story this information is used to let the character introduces herself. In addition, the information is also used to write a prompt for the AI tool to *generate an image* described in section c) of this chapter.

### ***b) Find a Meaningful Name for the Main Character***

A very characteristic attribute of a person is their name. Therefore, I searched for a suitable name. I document the process below. A companion question is whether this process of *naming* can be generalized to other data stories.

First, I combined the gender, year of birth and nationality/place of residence of the subject and searched for the most popular first names for girls in 1933, Germany.

1. Helga	11. Ruth	21. Christel	31. Karin
2. Ursula	12. Erika	22. Maria	32. Marion
3. Ingrid	13. Hildegard	23. Gertrud	33. Brigitte
4. Gisela	14. Ingeborg	24. Anneliese	34. Lisa
5. Gerda	15. Hannelore	25. Elfriede	35. Margarete
6. Ilse	16. Edith	26. Marianne	
7. Inge	17. Renate	27. Elisabeth	
8. Christa	18. Lieselotte	28. Rita	
9. Waltraud	19. Margot	29. Rosemarie	
10. Irmgard	20. Eva	30. Anna	

Helga is the female version of Helge and means ‘happy’, ‘healthy’ or ‘holy’. The meaning would fit well into the concept but it sounds antiquarian and is not popular nowadays.

Thus, I researched for popular names from the period the subject grew up and found a list of the most popular girls’ names in Germany from 1933–1945.

1. Emma	11. Mathilde	21. Wanda	31. Roberta
2. Lina	12. Friederike	22. Hilde	32. Alberta
3. Ella	13. Hedda	23. Karoline	33. Berta
4. Ida	14. Erna	24. Wilhelmine	34. Olga
5. Frieda	15. Ada	25. Amalie	35. Ottilie
6. Rosa	16. Irma	26. Edith	36. Senta
7. Karla	17. Adele	27. Hedwig	
8. Selma	18. Erika	28. Elvira	
9. Hermine	19. Henrike	29. Gerda	
10. Wilma	20. Minna	30. Thekla	

(<https://blog.beliebte-vornamen.de/2018/08/nazi-namen/>)

*Emma* is a female name that was popular during the time the subject was growing up (1933–1945) and is popular again today. Moreover, in Germany, the name is most popular in Mecklenburg-West Pomerania where the subject has her place of residence (<https://www.beliebte-vornamen.de/4513-emma.htm>).

For the last name, I thought about a name that is in harmony with the first name (Emma), that means of similar length, the reading feels easy, no tongue twister as well as first and last name do not rhyme. For inspiration I searched at a portal for names where last names are grouped after popularity visualized by word cloud (<https://www.namen-liste.de/nachnamen.html>). Moreover, I was looking for names that are more popular in Northern Germany as the protagonist comes from Pomerania. But last names such as “Eriksson” or “Petersson” sound too Scandinavian and I rejected those names from the list. I also looked for a name that is literary valuable, that means it excites some imagination or sound strong such as “Wolf”, “Winter” or “Goldberg”. The name also has to be historically neutral, so “Goldberg” is very common as a Jewish last name and because the religious affiliation was not included in the data, I removed that name from the list as well. My final condition was that it would be optimal to have a name that works as well in German as it does in English. So, “Winter” remained.

Finally, “*Emma Winter*” is the protagonist’s name.

### ***c) Generate Images for the Characters***

For the imagery and animation of the characters, I experimented with AI tools, such as *Midjourney*\* and *Leonardo.AI*\*\* to generate a digital image by writing a textual description (imagine prompt). Both Midjourney and Leonardo.AI use a *Stable Diffusion Model* (\* <https://www.midjourney.com>, \*\*<https://app.leonardo.ai/>).

*Stable Diffusion* is a deep learning model that is specialized in converting text to images and can also be used for ‘image-to-image translation’. For storytelling, this feature is relevant when similar or related images are needed in the same context. *Stable Diffusion* is a *Latent Diffusion Model*, i.e., a deep generative neural network [Rombach 2022]. ‘Diffusion’ describes the process of denoising a normally distributed variable or data distribution. “These models can be interpreted as an equally weighted sequence of denoising autoencoders which are trained to predict a denoised variant of their input” [Rombach 2022]. The term ‘stable’ refers to the enhancements of the Latent Diffusion Model which work on most consumer hardware devices with at least a modest GPU and is therefore no longer limited to cloud services (<https://stability.ai/blog/stablediffusion2-1-release7-dec-2022>). An example for the model architecture is shown in Fig. 16.



*Figure 16: LDM conditioned either via concatenation or by a more general cross-attention mechanism [Rombach 2022].*

At the end of this experimental section, I will answer the following questions: «How to generate good quality and suitable images with AI image generation?» and «How to receive a consistent identity and appearance of the character?»

The model has to be trained or needs an input image as reference. The order of textual fragments influences the extent to which they influence the final image, with words at the beginning having greater influence. Some tools also work with explicit weights.

The general structure of a prompt is as follows:

*/imagine/ <character name> <textual description of the person>*

## **Experiments with AI Image Generation**

I first tried out Midjourney because it is the most advanced AI tool for generating digital art. But the free trail is very limited and therefore, I switched to Leonardo.AI which is a very good alternative and gives me more freedom to experiment.

Prompt: Emma Winter A beautiful, semirealistic, 65 years old, overweight, educated woman, looks into the camera, with a friendly, independent, open minded personality, wearing glasses, platin-blond, long pixie hairstyle, she is a mother and was a kindergarten worker --no text --no letter

### *Results with Midjourney AI model*



For each of the following eight images, I used the same prompt: */imagine Emma Winter A beautiful semirealistic 65 years old, over-weight, highly educated woman with a friendly, strong, independent, open minded personality, she loves children, she is a mother and was a kindergarten worker, she has a fatty liver and an anxiety disorder no text --q 2 --v 4*

*Although I entered in the same prompt, very different facial expressions and characters were returned. They do not look more like gaming characters.*

*The second challenge was to generate a friendly facial expression.*

*The attribute anxiety disorder may have brought some darkness to the facial expression, so I deleted this attribute. I also slightly changed the prompt and experimented with weighting of different parts of the prompt.*

*Figure 17: AI-generated portraits by Midjourney AI*



*Examples of a semi-realistic image: Here I have been inspired for the long pixie haircut of my character which I think is more modern*

*Figure 18: AI-generated portraits by Midjourney AI*

According to these results, I looked for a more modern, friendly and realistic character which is more relatable and personable/likeable.

### ***Results with Leonardo.AI***



→ One of the first results with Leonardo.AI

Prompt: *Emma Winter A beautiful, friendly, semirealistic, 65 years old, overweight, educated woman with a friendly, independent, open minded personality, platin-blond, wearing glasses, long pixie hairstyle, she is a mother and was a kindergarten worker --no text --no letters*

*Figure 19: AI-generated portrait by Leonardo.AI*





Figure 20: Portraits by Leonardo.AI

Example of a photorealistic image of the same prompt (text description only differ in using the “*photorealistic*” instead of “*semi-realistic*”. The image is very realistic but overweight is not that visible. However, the overall impression is a good starting point.

#### **Image-to-Image Translation:**

In Leonardo.AI, I can give the model an input image as reference and it will produce variations. This method I can also use for generating images for different time steps, where the person changes in terms of age, weight or emotional expression for example.

#### ***Final image for the protagonist***

Prompt: *Emma Winter A beautiful, friendly, photorealistic, 65 years old, overweight, educated woman with a friendly, independent, open minded personality, platin-blond, long pixie hairstyle, wearing glasses, she is a mother and was a kindergarten worker --no text --no letters*

#### ***d) Create Variations of the Character***

The character changes her age, body weight and lifestyle over time on the one hand and on the other hand variation of the facial expression are needed to transport different emotions which is an important aspect of storytelling. Therefore, I created one general description of the character (general appearance and personality features) that keep stable over time and for each time step or different scenes there is a variation of the character, i.e., age, weight, emotions, clothing. Therefore, I experimented with image-to-image translations guided by a text prompt. A negative prompt embraces all elements or features should not be included, e.g., text elements.

Table 6: Character description in different time steps

	time step 1	time step 2	time step 3
General description	Emma Winter, semi/photorealistic, high quality, beautiful, woman, friendly, open-minded personality, looks into the camera, wearing glasses, platin-blond long pixie hair		
Negative prompt	text, letters		
Age	65	70	77
Weight	overweight	normal weight, losing weight	athletic, slim, normal weight
Emotional expression and special features	<u>Introduction</u> : positive, friendly, open-minded <u>Diagnosis</u> : fearful		proud, happy, living an active, happy and healthy life
Clothing	casual clothing	wearing sports clothing	wearing outdoor clothing (hiking)

***Character description for time step 1:***

*/imagine* Emma Winter, photorealistic, beautiful, overweight, 65 years old woman, overweight, friendly, open-minded personality, full body portrait, looks into the camera, wearing glasses, platin-blond long pixie haircut

***Changed description for time step 2:***

*/imagine* Emma Winter, photorealistic, beautiful, normal weight, 70 years old woman, overweight, large waist circumference, friendly, open-minded personality, full body portrait, looks into the camera, wearing glasses, platin-blond long pixie haircut, sports clothing

***Changed description for time step 3:***

*/imagine* Emma Winter, photorealistic, beautiful, athletic, slim, normal weight, 77 years old woman, overweight, large waist circumference, friendly, open-minded personality, full body portrait, looks into the camera, wearing glasses, platin-blond long pixie haircut, outdoor clothing



### *e) Visualizing Lifestyle Change with different Clothing and Emotions*

To visualize the life style change of our protagonist, I came up with the idea of having the character wear sporty clothes.

With an image-to-image translation, I tried to keep the visual identity of the character and only change the age and the cloths to more athletic by adding “wearing sports clothing” or more specific e.g. “blue-white jogging jacket” to the prompt. But the model produces similar clothing as in the original image. Therefore, I first generated the clothing in an appropriated perspective with the prompt “*woman in blue-white jogging jacket frontal*”, then I created a collage out of both images and entered this image back as input to the model with a prompt that changes the age to 70. Then I selected the image that matches best with the original image and has a strong-positive facial expression as well.

With *image-to-image translation* I have generated variations by for:

a) changing emotional expression



***Introduction*** (seed-image)  
*Positive emotion: joyful, friendly  
(likeable/personable)*



***Emotional reaction to diagnosis***  
*emotion: fear, worried*  
*Post-processing: Some features need to  
be adjusted to make them look stable over  
time, e.g. the color of hair or skin*

b) change of her physical appearance by getting older in the next time steps and losing weight simultaneously



*70 years old and normal weight*



*77 years old and athletic/slim*

Stable Diffusion Model



Protagonist



Stable Diffusion Model



Clothing



Adobe Photoshop



Collage

Stable Diffusion Model

Prompt: *Emma Winter A friendly, beautiful, photorealistic, 70 years old, normal weight, educated woman with an independent, open minded personality, platin-blond long pixie haircut, wearing glasses, she is a mother and was a kindergarten worker*



*70 years old, normal weight with sports clothing*



## *“Hero end pose”*



### ***Step 1 Image Generation***

Prompt: *athletic woman in outdoor clothing jack wolfskin frontal (define a clothing style by a fashion brand, Image errors must be post-processed)*

### ***Step 2a: Collage***



### ***Step 2b Image-to-Image Translation***

Prompt: *Emma Winter A friendly, beautiful, photorealistic, 75 years old, athletic, woman in Outdoor clothing jack wolfskin frontal (Result: a very different head is generated)*

### ***Step 3 Postprocessing***

*Due to the problems in Step 2b (when the head of the person is too small Image-to-image translation generates a very different head), I used the collage 2a. The image can be improved by postprocessing, e.g., correction of depth of field, other image errors*

## *The Doctor*

Can be added as a real photograph of a physician (liver expert, internist) that is animated by AI. The same is possible with a computer-generated image when a real physician is not available.



*The specification “wearing” is important because the prompt “50 years old doctor glasses” leads to a strong focus on the subject of “glasses”*



*Prompt: 50 years old male doctor wearing glasses*



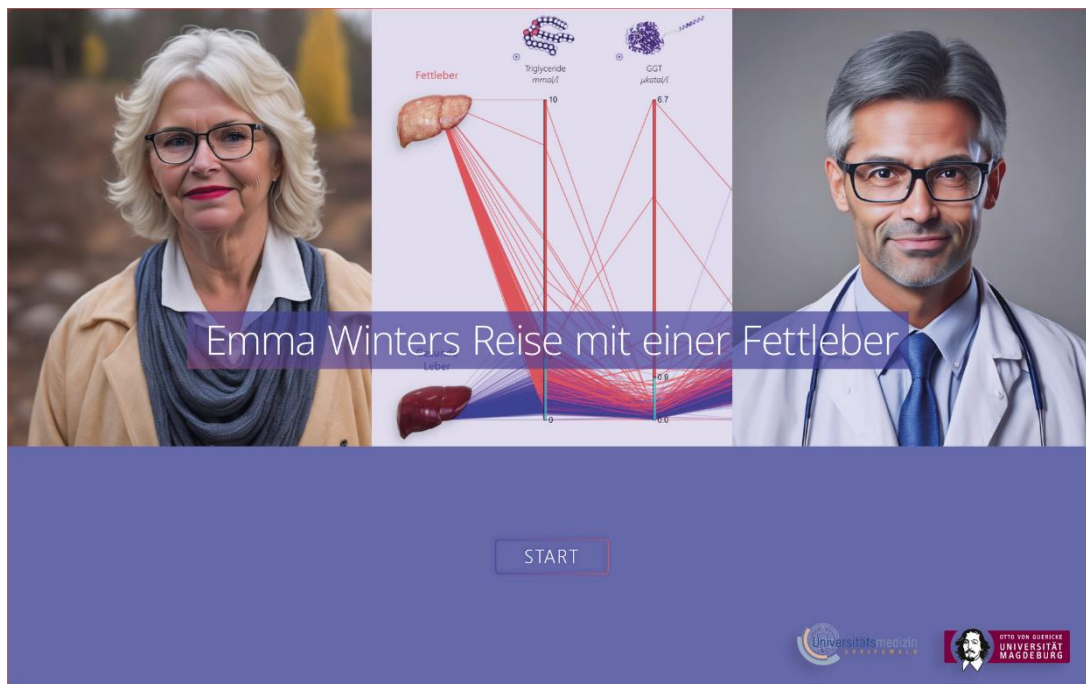
*Original image*



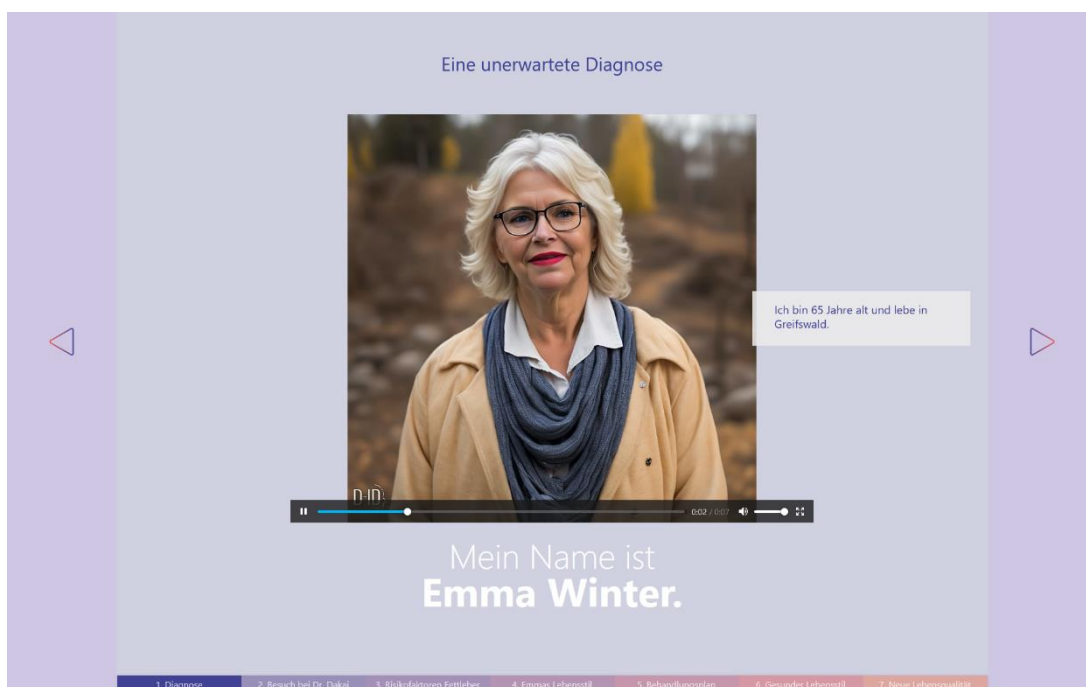
*Image-to-Image Translation:  
using the original image and a variation  
of the original prompt*

*Prompt: 60 years old male doctor  
wearing glasses*

## Realization: Examples for the Use of the Character Images



*Collage for the Menu of the Story Prototype*

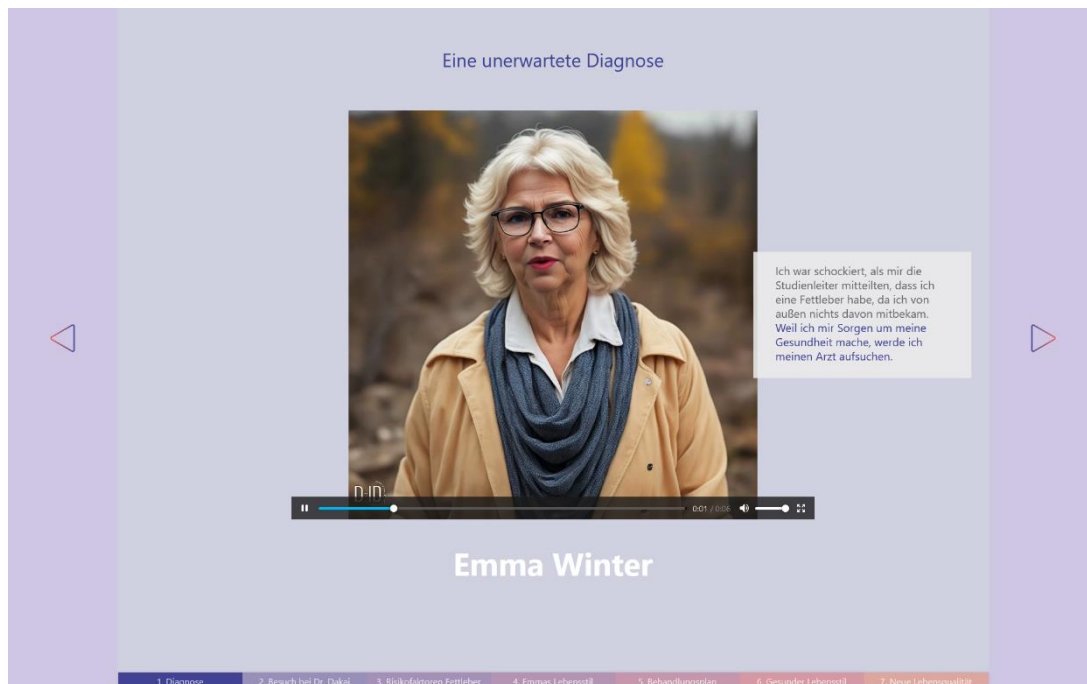


*The protagonist introducing herself using AI Video generation*

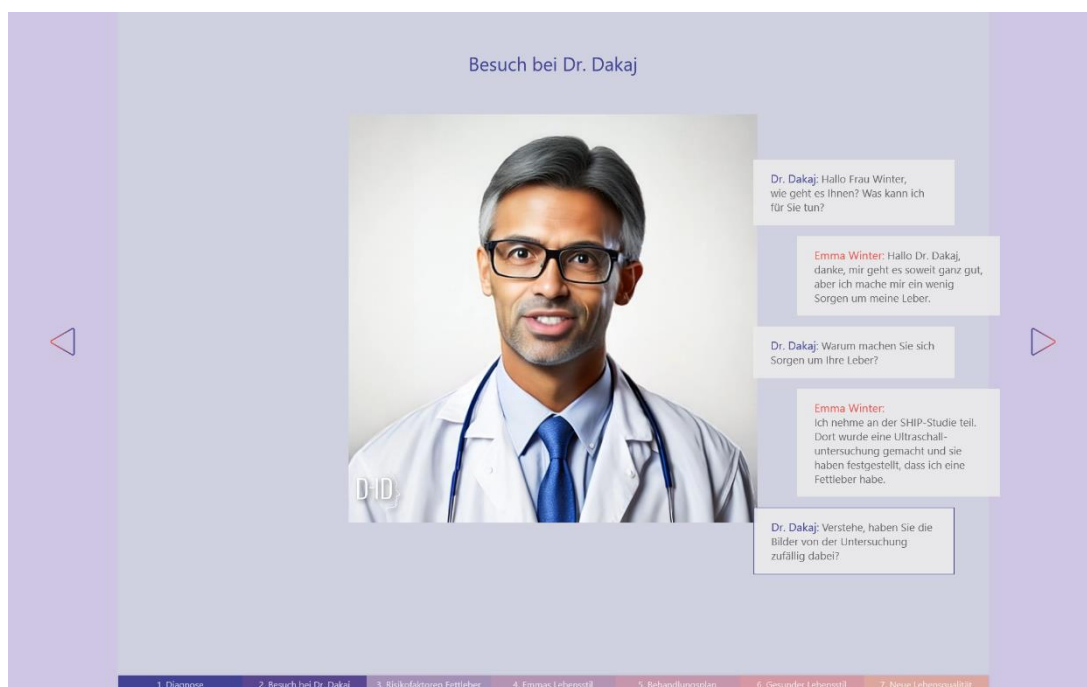
*Tool: D-ID (<https://www.d-id.com/>)*



## Realization



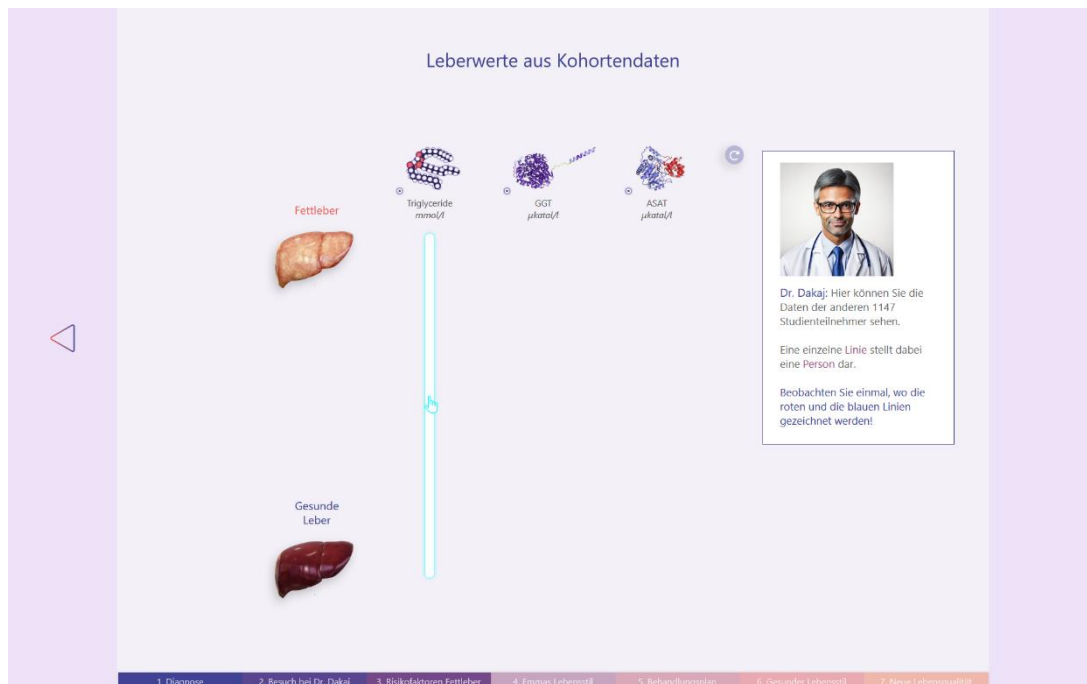
*The emotions of the protagonist are expressed even better in the video*



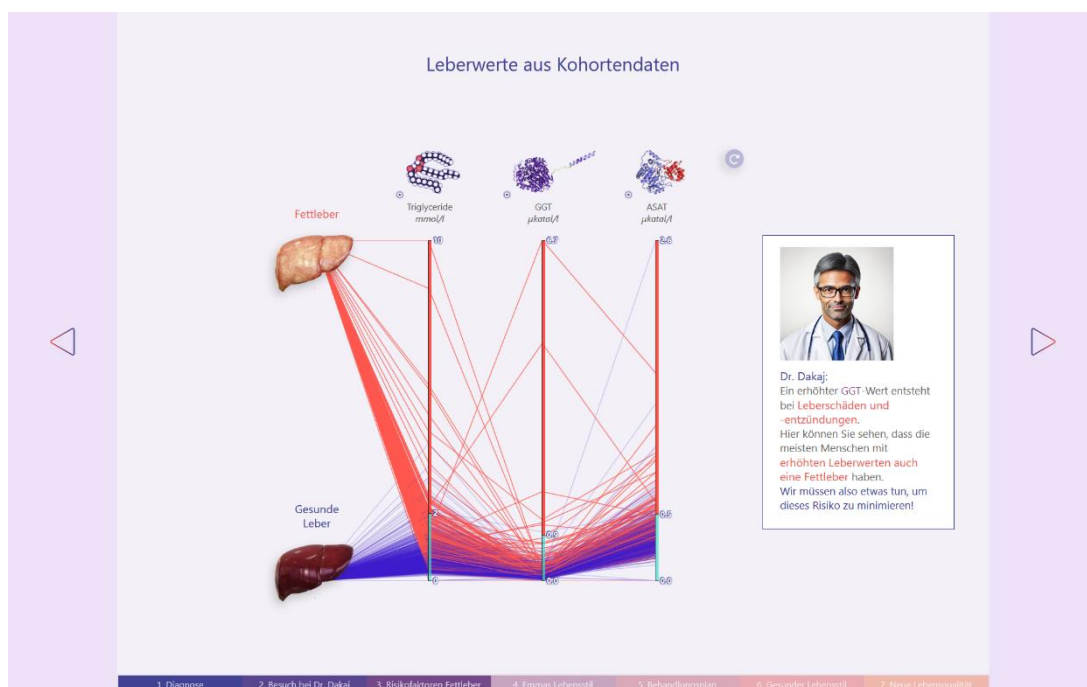
*A dialog design with alternating video sequences*

*Tool: D-ID (<https://www.d-id.com/>)*

## Realization



A dialog window introduces the diagram: The doctor as a thumbnail\*

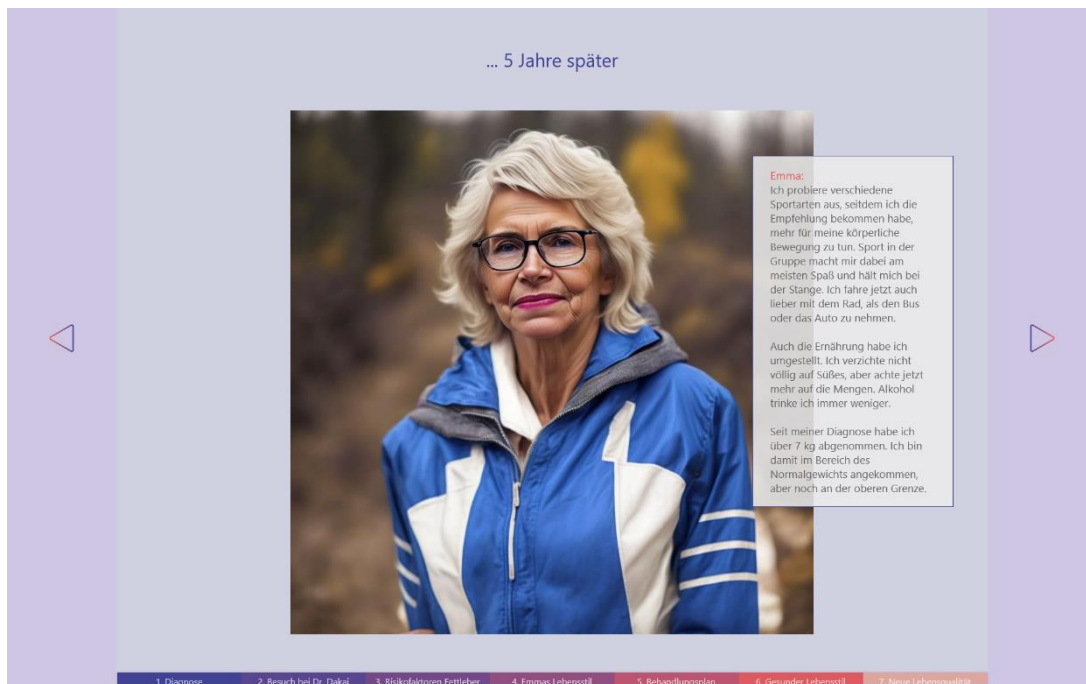


When the diagram has built up, the doctor draws a conclusion

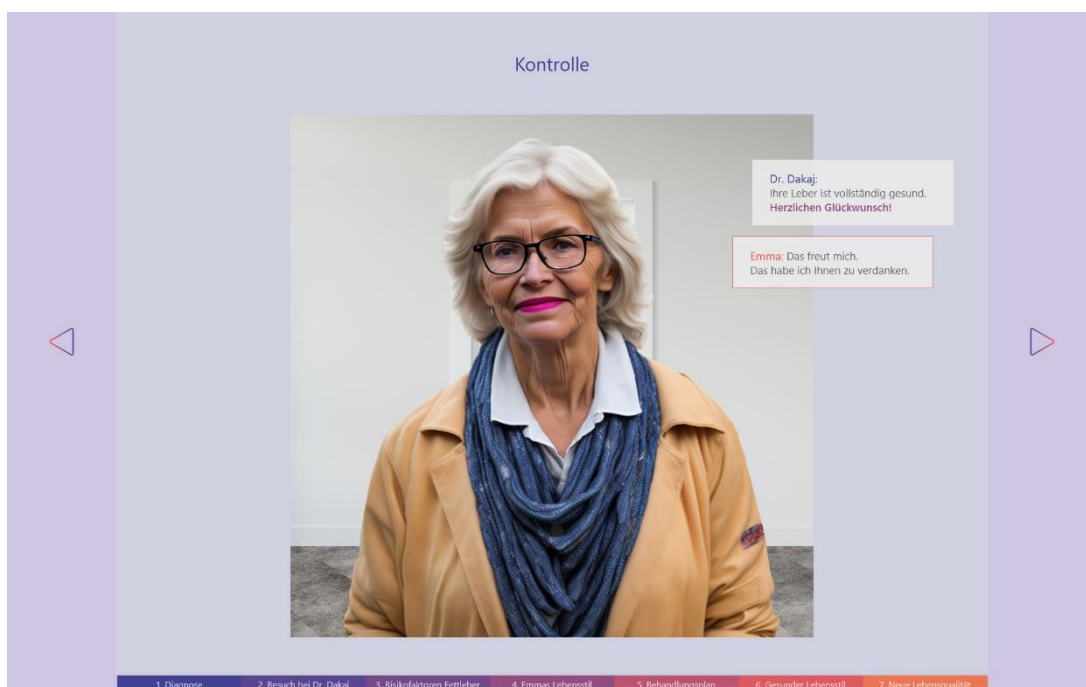
\*thumbnail = small representation



## Realization

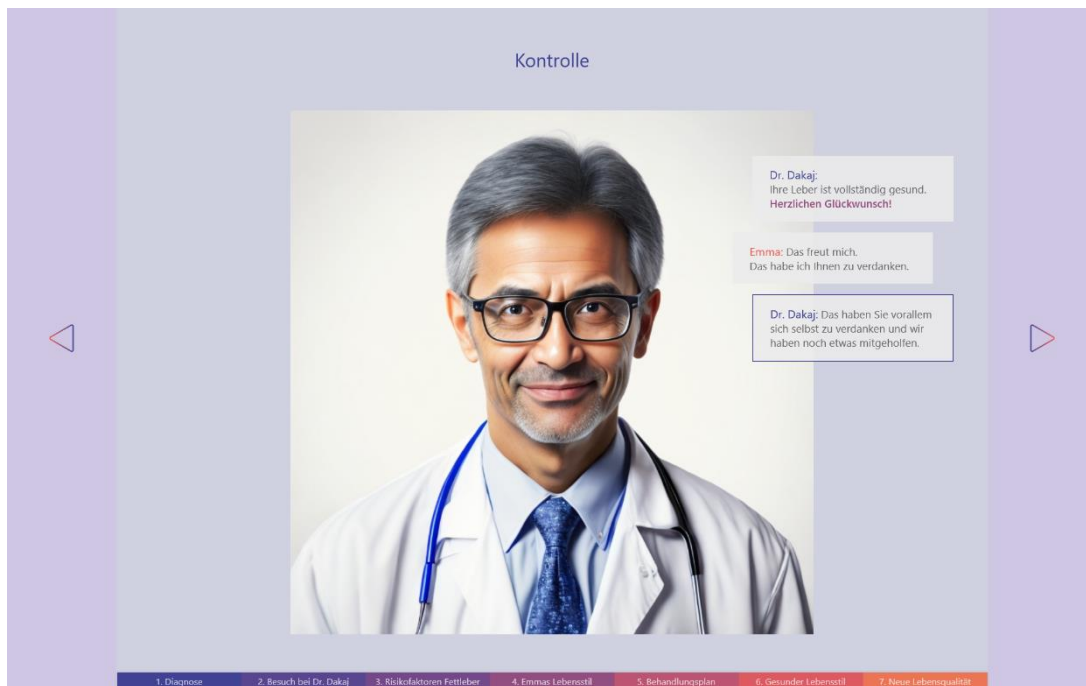


*The protagonist five years later where she is 70 years old and lost weight*

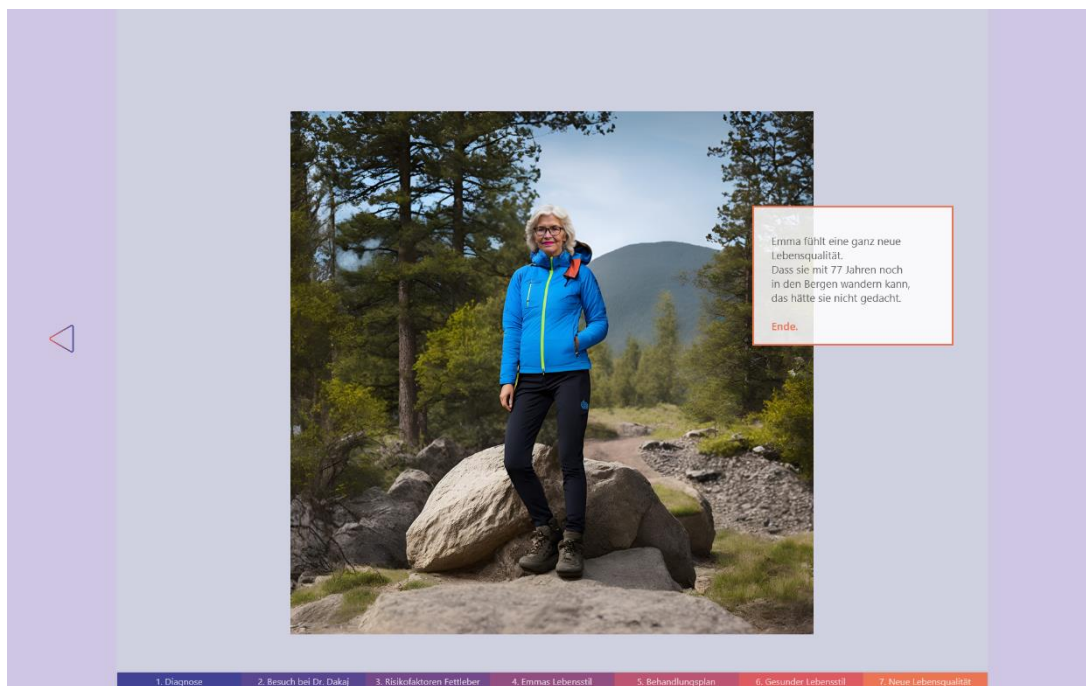


*When the protagonist is 70 years old, she has reached a normal weight and a healthy waist-to-height ratio as well.*

## Realization



*The doctor has aged as well, with important facial features preserved.*



*The story ends with a hero shot.*

## 4.4 Story Design

The story concept combines a Freytag's Pyramid story pattern, AI-generated protagonists (based on cohort data), incremental construction (interaction) and gamification elements to enhance User Engagement.

### Story pattern

The story design combines the 'Freytag's Pyramid' with the idea of that the protagonist is a 'hero'.

1. Introducing the protagonist, diagnose, data background
2. Introducing the doctor, information and relevance of the health problem
3. Risk factors with data visualization (Parallel Coordinates introducing)
4. Life circumstances of the patient
5. Treatment plan (general areas of lifestyle to concrete behavioral changes)
6. Lifestyle changes (realization of behavioral changes by the protagonist)
7. Resolution: Negative diagnosis, protagonist living a healthy and active life

### Story summary

PART 1 (Introduction): Emma Winter, a 65-years-old woman, was diagnosed with fatty liver via ultrasound imagery during the examinations in Study of Health in Pomerania (SHIP). A brief information about SHIP and cohort studies is narrated by the protagonist. After the diagnosis, she expresses concerns about her health and decides to consult her doctor.

PART 2: Emma visits Doctor Dakaj. Emma tells her doctor about the concerns she has about her liver health. The doctor asks for documents from the SHIP study (ultrasound images). He explains what is shown on the ultrasound image and what a non-alcoholic fatty liver is by a 3D model. He warns the patient about serious consequential liver diseases that can develop from fatty liver, such as liver cirrhosis and liver cancer. The patient is asked to stand on a body scale to determine her weight, and a nurse takes a blood sample from her.

PART 3: Parallel Coordinates are motivated by the doctor as an overview technique for visualizing relations in her data and to learn about risk factors. Parallel Coordinates are introduced by a thermometer metaphor and a task. After that the results from Emma's blood test are shown. Three liver function values are displayed in PC plot and discussed by the doctor. The test result shows that her Triglycerides are too high but other liver function values that show liver

inflammation are not yet out of normal range. Then the doctor shows her the main risk factors for fatty by a PC plot with cohort data. [The increase of fatty liver diseases related to gaining weight is visualized by a time-oriented PC plot.]

PART 4: The protagonist talks about her life (circumstances).

PART 5: A treatment plan is created with a quiz form where the user is asked to decide in which areas of life the protagonist needs to change her lifestyle and then decide on specific behaviors, such as reducing carbohydrates and doing a sport she enjoys 2-3x per week

PART 6: She chooses concrete actions to change her daily life.

As consequence she loses weight.

PART 7: When she is 77 years old, her liver is again screened by ultrasonic imagery. She is diagnosed negatively. Her liver is much healthier now. Liver function values are in normal range or have decreased. The doctor congratulates her for the successes. She enjoys her life and is proud to be able to live such an active life (hiking scene).

Table 7: Planning the story

	Content and Visualization
Menu	<ol style="list-style-type: none"> <li>1. story title,</li> <li>2. navigation buttons,</li> <li>3. important visual elements used in the story as collage (characters, disease-related body part, data visualization, symbols for lifestyle change, e.g. sports clothing)</li> <li>4. logos</li> </ol>
<b>PART 1: Introduction: “DIAGNOSIS”</b>	
Introducing the protagonist and its diagnosis	An AI generated video sequence with the protagonist introduces herself: name, age, occupation, family status and that she participated in SHIP
Excursus: SHIP (Short introduction)	<ol style="list-style-type: none"> <li>1. What is SHIP?</li> <li>2. What is a cohort study?</li> </ol>
Relevance: Frequencies in (a) gender and (b) age	Data visualization of frequencies: (a) pie chart), (b) age (Isotype charts with pictographs of human figures)
Emma is concerned about her health	The image of the protagonist is slightly changed to a concerned facial expression. In a short video sequence, she says that she wants to see her doctor.

PART 2: “VISITING MY DOCTOR”	
Dialog with the doctor	Introducing the doctor
Game element: Ultrasound images	The user is asked to choose the right image the doctor is asking for: Ultrasound-images of fatty liver
What is Fatty Liver? (definition, visual appearance in an ultrasound image)	While the doctor talks about how to read the ultrasound images, he explains the definition of a fatty liver and its characteristics compared to a normal liver
Why is a healthy liver so important? What are the functions of the liver?	To emphasize the importance of a healthy liver, the user can explore an aesthetic 3D model of the organ and the most important tasks of the liver by hovering over icons. A tooltip provides further information.
Story peak: What are consequential diseases?	The models change in: (a) color, (b) surface texture and (c) volume Liver stages: (1) fatty liver*, (2) hepatitis (inflammation) **, (3) fibrosis (scarring), (4) cirrhosis**, (5) liver cancer**, Patient’s emotional reaction: * disgust, ** fear
Reassuring words and preparation for examination	The doctor tries to reassure his patient (liver regeneration). He wants to examine the state that he wants determine her body weight and examine her liver values by a blood test.
Examinations:	The doctor instructs the patient to go on a body scale to weight her and then to a nurse to draw her blood. The blood sample is send to the lab. The user can progress the story or go back to the doctor for more information about the examination steps.
PART 3: “RISK FACTORS OF FATTY LIVER”	
Intro dialog	The doctor would like to discuss the risk factors with his patient (transition and motivation why to use parallel coordinates: overview about several factors, comparing with other subjects having fatty liver, what are common characteristics?)
Introductory example to Parallel Coordinates (“Learn-by-doing”-Game)	Game element: Read the notice of the doctor for Emma’s age, weight and height. A “thermometer” metaphor as visual cue how to read the axis (low values on the bottom, high values on the top). Attributes: age, weight, height, BMI The user has to set the correct points on the axis of a Parallel Coordinates plot. After selecting a correct point, a line is drawn between to axes.

	The BMI axes introduces another concepts of reading the scale, with a differently colored sections (normal range and too low or too high values)
Examine liver function values	Parallel Coordinates plot to examine liver function values: (1) Triglycerides, (2) GGT, (3) ALAT using incremental construction
Cohort data for the liver function values	A single time step for the cohort data: Triglycerides, GGT, ALAT (protagonist highlighted). Dialog windows: (1) start: explanation how to interpret the data (that a single polyline represents a person), question what the user notice about the distribution of the colored lines for fatty and non-fatty liver participants (2) end: review/resolution about the presented data
Parallel Coordinates Plot of risk factors for fatty liver (Incremental construction, first time step, four axes), cohort data + protagonist highlighted	A single time step of cohort data for Waist-to-height Ratio, BMI, Alcohol consumption, Diabetes Mellitus Type 2, categorical data (portion are shown on the axis), Highlighting Emma Winter as a path at the end (fulfilled 3 out of 4 risk factors)
Physical Activity (two time steps)	time-oriented PC plot shows the impact doing exercises on recover from fatty liver
<b>PART 4: “EMMA’S LIFE”</b>	
Protagonist’s life circumstances/ life style	Video sequence of the patient talking about her life circumstances <ul style="list-style-type: none"> <li>- kindergarten teacher, was very active during her working day, but now she is retired</li> <li>- loves to be with her family and friends, likes to cook and bake, they like to invite and entertain each other</li> <li>- does already some sports, namely Nordic walking with a group once a week</li> </ul>
<b>PART 5: “TREATMENT PLAN”</b>	
Reflection task: The users decide the realization of the treatment plan in two steps  Design principle: “From general to concrete actions”	The users shall reflect about what they have learnt about risk factors for fatty liver, and what healthy and unhealthy behaviors of the protagonist are.  <u>Interaction task 1:</u> The user has to decide what the parts of her treatment plan are, for example “increase physical activity”  <u>Interaction task 2:</u> The user has to find concrete actions to realise the treatment plan in everyday situations, for example in resp. to physical

	activity “ride the bike to visit friends”, “go to the local indoor swimming pool ones a week for one hour”) (inspired by Cognitive behavioral therapy (CBT))
<b>PART 6: “LIFESTYLE CHANGE”</b>	
Five years later the protagonist has established a healthier lifestyle	Emma reports the success of losing weight and how she did change her lifestyle
<b>PART 7: “NEW LIFE QUALITY”</b>	
Resolution: Examination in two steps (1) Blood test (2) Ultrasound imagery → Recovery and Motivation  Hero end pose	Protagonist visits her doctor again. (1) PC plot provides an overview about her liver function values over three time steps → Dialog: Motivation to keep on. (2) Her liver is screened by ultrasound → Congratulations from her doctor for recovery and emphasis of her ‘self-efficacy’ End: Protagonist’s reaction (77) is glad and proud about her recovery and enjoys a new life quality (a hero scene, e.g. being active in nature, climbing on the top of a mountain)

*The texts for the dialogues are inspired by the conversations between doctors and patients having fatty liver in the series “The nutrition docs” (NDR).*

## Storyboard

To organize the story and sketch first ideas and variations I used the storyboard technique. Although, the origin of storyboards lies in the field of filmmaking, this technique is applied to other areas, such as Web-based User Interface Design. Storyboards have the strength to reduce “complexity in uncertain environments and creating an early visualization” of systems [Rasheed et al. 2019].

In comparison to a movie a website has different UI elements which are also part of the storyboard. These elements can be subdivided into

- (1) *Informational UI Elements* (Label, Tool Tip, Icon, Pictographs, Progress Bar, Notification, Message Box),
- (2) *Navigational UI Elements* (Back and Forward Buttons, Start Button, Links),
- (3) *Input Control UI Elements* (e.g. User Input for answering a question, Checkbox, Radio Button, Text fields). A *scene* here is basically defined as an “user-inter-actable view” [Rasheed et al. 2019] or “screen layout” in a particular state [Landay & Myers 1996].

Another UI specific problem about sketching storyboards is how to display interactive elements and the specification “how the screen changed in response to user actions” [Landay & Myers 1996]. The authors propose an arrow technique to visualize the linking between UI elements.

In summary a storyboard contains the scene and annotations:

- Number of the scene
- Scenario title
- Simple sketch of a ‘scene’ in comic style
- Protagonists by name
- Short description
- Audio/Text
- UI Elements, Animation and Interaction

Based on this research, I created a storyboard template:

The storyboard template consists of two pages. The left page contains six storyboard cards arranged in a 3x2 grid. Each card has a circular icon at the top left, a rectangular sketch area, and three horizontal lines for text below. The right page contains two larger storyboard cards, one above the other, each with a circular icon at the top left, a large rectangular sketch area, and three horizontal lines for text below.

*See Appendix for sketches*



## **Gamification with interaction-directed Story Progress**

*Example in part 2: Visiting Doctor Dakaj: “Ultrasound Image”*

The user has to select the correct image when the doctor asks for “images of the examination”. Here the user had read/listen to a dialog and must combine information from the both the doctor and patient.

Combination task: Emma says she had done an ultrasound examination, Dr. Dakaj asks for images of the examination; Solution: ultrasound image

Action cue: “Look into Emma’s bag and give the image to the doctor”

Elements: A bag with MRI image of a brain, Picture of the last holiday, ultrasound images of fatty liver

Feedback:

- “positive”: The story progresses, rewards
- “negative”: Thoughts of the patient (humorous)

*Example in part 2: “Examining Liver Function values”:*

1. “Take blood from the patient”, 2. “Give the sample to the lab”,
3. “Bring the result to the doctor”

Decision task: “Is the value in good range?” Answer: “Yes” or “No”

*Example in part 3: “Introductory example to parallel coordinates”*

This interactive puzzle should provide to learn how to read parallel coordinates. The game is integrated within the story. The user is asked to help the doctor to order the information he received from the patient. The data is noted on a bracket folder that I have seen doctors use in hospitals. The list contains simple attributes such as age, weight, height and calculated BMI. The user is then asked to click at the button associated with each value. If the user is correct, then the button is highlighted with a different color than the others, and when the button is clicked, a connecting line segment is drawn from the patient’s name on the clip folder to the point on the axes. For the other points the polyline is continued accordingly to a Parallel Coordinates plot.

*Example in part 5: “Treatment plan”:* Users must click check boxes to create a treatment plan for the patient.

## 4.5. Design Choices for Parallel Coordinates

In the topic of risk factors, you have values that are good, bad or with metrical values something in between. So more bad values could be on the top and good or healthy values on the bottom. Then the subjects with a certain disease are colored differently than the persons without the disease, most of the first group with fatty liver is on the upper range of the axis.

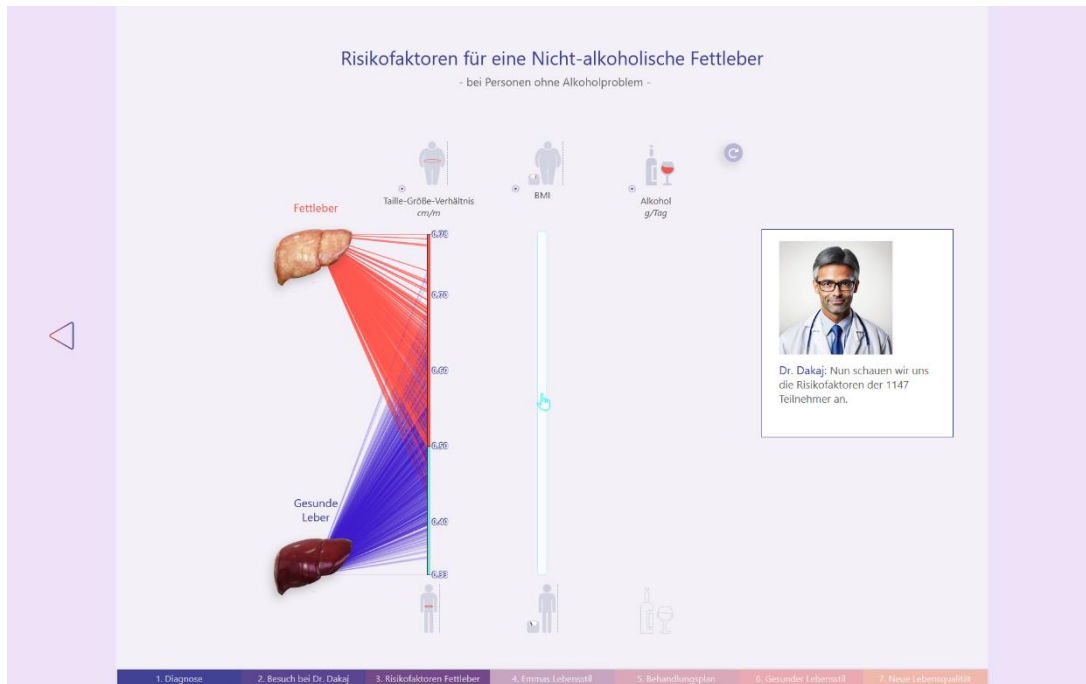
I implemented Parallel Coordinate plots with different complexity levels where the complexity is rising within the story progress.

- (1) Liver function values for at least the protagonist  
(a single time step)
- (2) Liver function values the cohort  
(a single time step)
- (3) Risk factors for the cohort  
(a single time step)
- (4) Liver function values for at least the protagonist  
(three time steps)

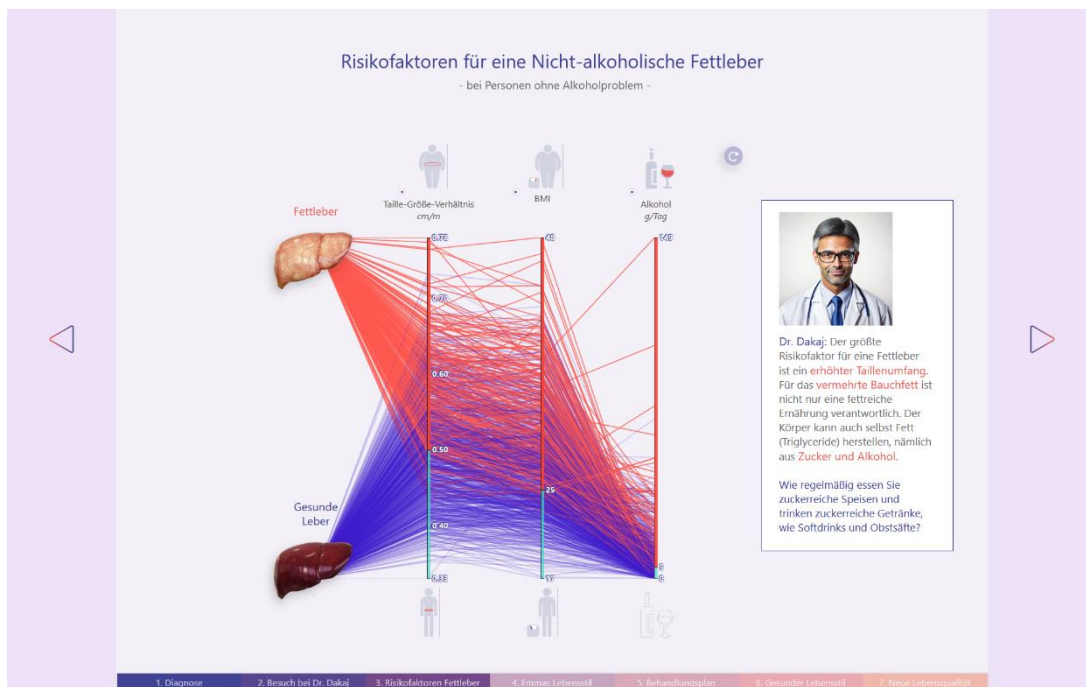
### Incremental Construction

I include the *reader-driven approach* [Segel & Heer 2010] with the incremental construction technique, interactive annotations, and variations of dialogues. However, the *author-driven* portion predominates due to the primary goal to foster comprehension of the complex nature of the underlying data.

The user can build up the Parallel Coordinates step by step (*axis by axis*). The first axis is displayed when the user clicks on a glowing button that is at the same position and has the same length as the axis. The glowing effect goes from ice blue to white to attract the user's attention and should be a visual and intuitive hint for interaction.



*Incremental construction: after the first interaction*

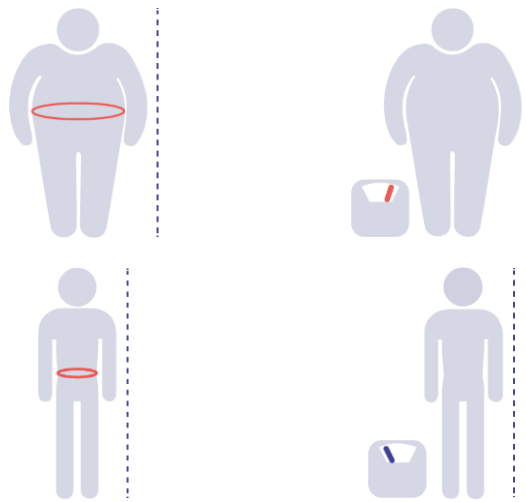


*Incremental construction: after the third interaction finishing the diagram construction, at the same time the explanatory dialogue is displayed*

## Annotation & Icon

A flat and simple design of icons captures the meaning. Additional information and explanations, e.g., about the variables is hidden by these interactive icons to reduce the complexity. This concept is intended to increase the user's joy of exploration and thus enhance User Engagement.

The axis has not just a label but a symbol that visualize the variable. The risk factor PC (3) has two icons – one above for the high value and one at the bottom for the lower value. For example, I took the highest waist-to-height ratio that could be found in the data set and map onto the waist-to-height ratio of a pictographic figure, the same I have done for the leanest participant in the data set.

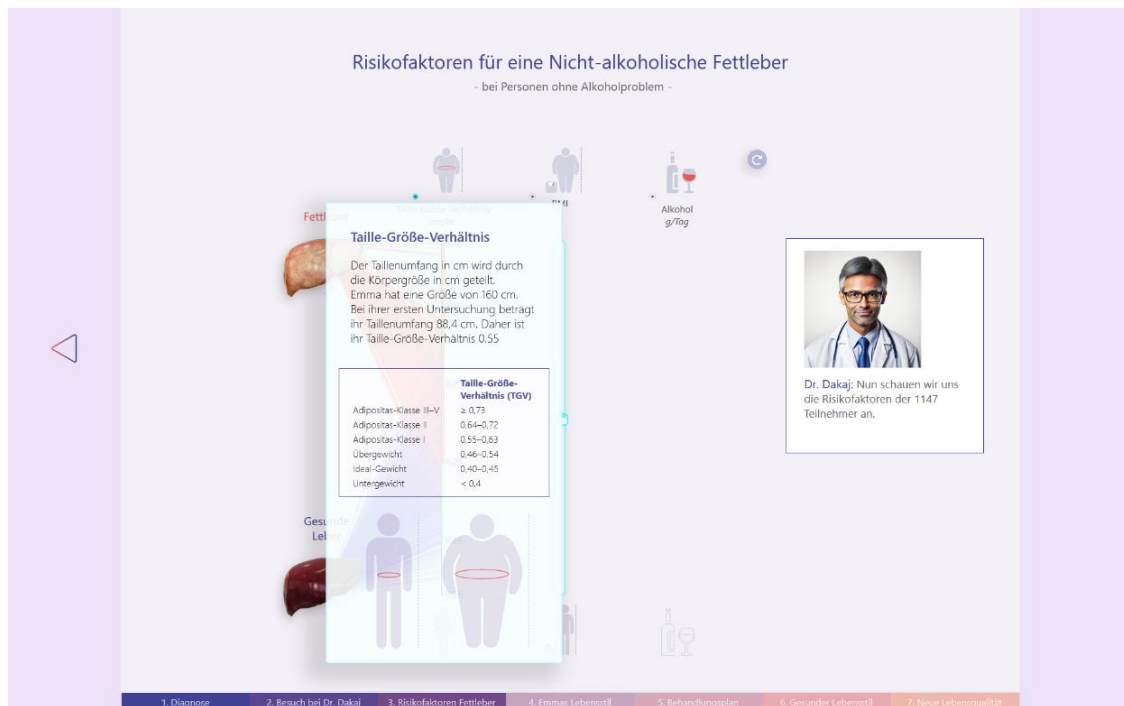


*Figure 21: Icons for highest (top) and lowest (bottom) value for Waist-to-height ratio (left) and BMI (right), [own work]*

## Tooltips

For the design of the tooltips, I was inspired by the metaphor of a glass plate on which information is displayed, but you can also glaze through it. Tooltips provide information or a definition for different variables, e.g., liver values or risk factors. Here the protagonist still plays a role as her data is used as a calculation example. Then the information is put in the context of the cohort data by a table that depict different groups.

However, very important information that is needed to transport the message or tell the story is repeated within the dialogues.



*Tooltip for each axis, hidden behind the pictographs at the top*

## Color schema

The liver should be displayed realistically. So we have a yellowish color for the fatty liver and dark red for the healthy liver. As mentioned in the introduction, colors are better to discriminate if they are different in hue, saturation and brightness. I thought of a color that is complementary to yellowish as main color, that is violet. Then I used the rule to distinguish between subjects having fatty liver with a signal reddish color and in contrast a transparent coolish blue-violett for individual with a normal liver.

Red is a warning color. So I used a bright red that stands out from my main color spectrum as first accent color, e.g. for the lines of the subjects with fatty liver and text elements. On the basis of these two major colors, I created a color scale for the navigation bar and other interaction elements that shall help to support the direction of view. View direction here refers to the way the user goes through a layout, e.g., a web page and can be tracked by an eye-tracking system [Pan 2004].

## 5 Implementation

The web application has been implemented with HTML5, CSS and JavaScript. For the data visualization and axes I used the D3.js a JavaScript library for manipulating documents based on data ([d3js.org/](https://d3js.org/)).

For data pre-processing I worked with Microsoft Excel and Libre Office Calc. In GitHub the final data set was uploaded to make it available to the d3 library.

On turbosquid website ([www.turbosquid.com/](https://www.turbosquid.com/)) I found a 3D model of a liver with cirrhosis. I edited the model with Blender in order to get variations for a healthy liver and fatty liver. The models were uploaded to the web-based platform Glitch ([glitch.com/](https://glitch.com/)) and the link was inserted within the html component model-viewer ([modelviewer.dev/](https://modelviewer.dev/)).

### 5.1 Web-based Concept

The website concept reminds of an interactive slide shows, where each slide is an HTML document. The different html documents are connected with forward and backward buttons and an interactive navigation bar at the bottom. Some subpages belong to a “excursus” and differ from the others by a missing navigation bar, a different background color and style of navigation buttons.

Some web pages do not have a forward button, when the user is asked to advance the story by solving puzzles, interacting accordingly by clicking the correct buttons/links. In this case, the concept adapts elements of a point-and-click adventure game.

### 5.2 Methods Documentation

*function drawlines(dimensions, color1, color2)* plots the PC, where parameter *dimensions* is a given array of selected variables from the data set, and *color1* is the color code for the subjects with fatty liver and *color2* encodes subjects with healthy liver.

*document.querySelector()* method returns the first HTML element within the document that matches the specified selector, or group of selectors.

*addEventListener()* method is used to implement the on-click events for the incremental construction of the Parallel Coordinates plots.

*function add(dom,ticks,link)* builds the axes

*function addreference(dom,ran,link,color)* adds different colored sections to the axes to visualize the normal range and therapeutic/critical range.

## 6 Evaluation

The goal of the evaluation is to test my techniques with a small sample of people to determine, first, if the prototype implementation is appropriate for the goals of the requirements, and second, to get qualitative feedback from users to improve the prototype.

However, “measuring the quality or success of visualizations or visualization techniques with respect to specific communication goals is a complex task.” [Böttinger 2020]. In the book “Foundations of Data Visualization” the authors provide a broad range of information about empirical studies in the science communication field of evaluating visualizations or visualization techniques. First, *design choices*, e.g., for the visual design and secondly, *deployed techniques* such as story pattern, animation, interaction techniques, have to be evaluated with a focus on (a) *effectiveness* and (b) *user engagement*. Measuring engagement is the most difficult part [Böttinger 2020].

### 6.1 Early feedback

Human-centered design uses the technique of iterative cycles of user feedback. In order to get this *early feedback* in terms of perception and understanding, I regularly surveyed colleagues and friends with and without experience with the SHIP data set. Either in the form of a short presentation of the research questions and with the state of the prototype followed by a discussion or a dialog-based feedback while going through the prototype. Later, I also pretested the interaction by letting colleagues and friends solve the tasks without giving detailed instruction.

### 6.2 User Study

I evaluated different complex variants of the data visualization within the same story framework and examine the impact of the ‘access techniques’ for new users of Parallel Coordinates. The story is about risk factors of *non-alcoholic fatty liver* with a female patient-protagonist and a male doctor-protagonist. The story will contain tasks. I got qualitative feedback by think aloud method and interviews of the test persons at the end.

It takes a person to go through the whole story ranged from 20 minutes to about an hour. The average was 35 minutes. The older subjects (62 and 66 years old) took the longest. They took many breaks to reflect on their own relationship to the topic, especially their own risk factors and doctor visits. One study participant is undergoing treatment for fatty liver and matched her own knowledge with the content of the story.

After finishing the stories, I let them fill out a questionnaire with a number of questions to retrieve what they have memorized about the topic, how well the data visualization (Parallel Coordinates) was understood and how well they felt engaged.

I examine three distinct domains to investigate the impact of the techniques used on (I) *Memorability*, (II) *Comprehension* and (III) *User Engagement*.

See questionnaire in Appendix (German version)

### **(I) Memorability**

To achieve a realistic situation, study participants were not informed at the beginning of the study that they would be asked to answer questions in a questionnaire about what they remember.

The participants perform the following tasks with multiple choice:

1. What is the name of the protagonist/patient?
2. Which medical examination technique is used to diagnose fatty liver?
3. What are characteristics of a fatty liver?
4. What are (avoidable) risk factors for fatty liver?
5. What are consequential diseases of fatty liver?
6. Which functions of the liver do you remember?
7. What are important behaviors to make healthy lifestyle changes?

### **(II) Comprehension**

- (a) I asked the participants to *think aloud* when they walk through the story, especially when they have any kind of *difficulties* or *positive experiences* as well.
- (b) Was *help* needed to solve tasks within the story? I sat next to the subjects and *observed* their interaction. If I noticed that they had problems or were stuck at a point, I did not help immediately. Some tried different options to solve a task or interact with an element in different ways, and most eventually found a solution.
- (c) The participants perform the following tasks in the questionnaire:



1. The user gets the PC plot of liver function values they already have seen in part 3 of the story: Which values are in normal range and which are not?
2. What does the line and its color in a PC plot means?

### (III) User Engagement

A scale-based questionnaire (after Amini 2018) was used to evaluate the five factors that have an impact on the User Engagement, and Interaction in addition:

- (A) Affective involvement (3 Items)
- (B) Enjoyment (2 Items, one open question)
- (C) Aesthetics (3 Items)
- (D) Focused attention (3 Items)
- (E) Cognitive involvement (4 Items)
- (F) Interaction (3 Items)

### Experimental setup

I tested my prototype between May 31 and June 4, 2023 with ten participants in the same room and on similar screens. Five test subjects during the “Lange Nacht der Wissenschaften 2023”. The web-based prototype was displayed using the Firefox browser. In one case I had to use the Edge browser. All test subjects saw the same *storyline*. I only tested different solutions of the *time-oriented parallel coordinates*, where each test subject sees only a single version.

Two versions of Temporal Parallel Coordinates:

<i>Variant 1</i>	<i>Variant 2</i>
Time-oriented PC build up by each variable and each time step (one axis per time step), nine interaction steps	Time-oriented PC build up by each time step (three axis per time step), 3 interactions steps



## 7 Results & Discussion

Although this user study was small, I have learnt a lot especially from observing users during the interaction, for example moving the mouse cursor, emotional reactions, comments and body language. I noticed that some users sit more straight when they have to solve a gamification task.

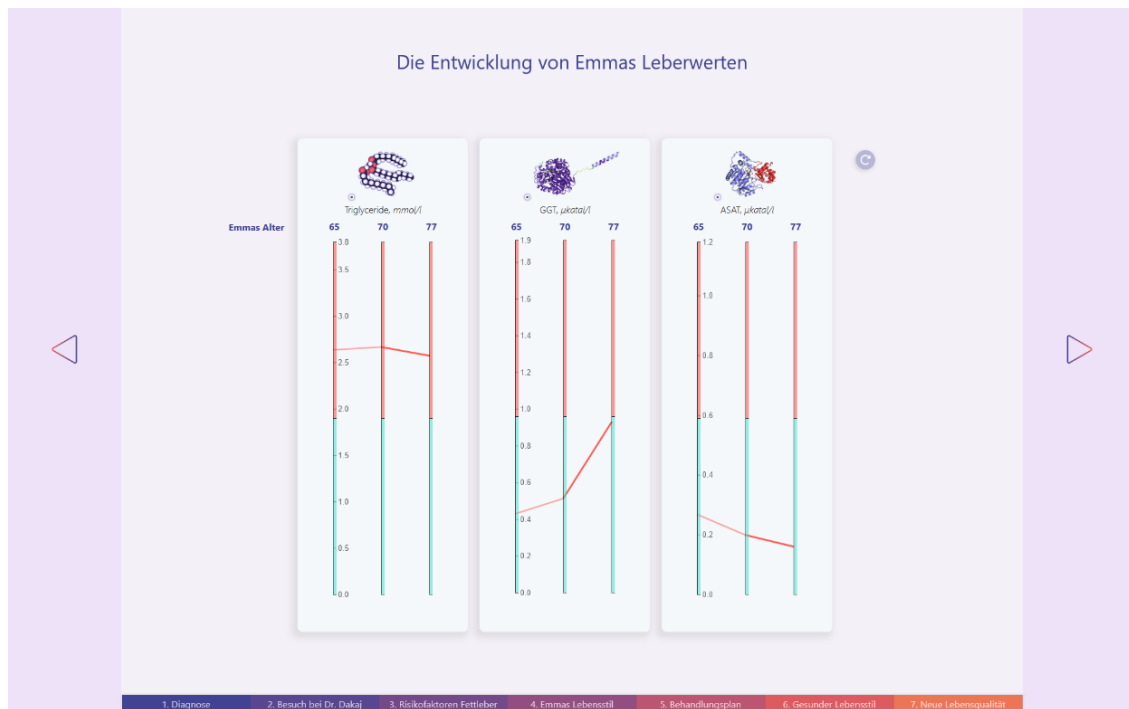
Getting qualitative feedback by the early feedback, the think aloud method and interviews of participants was a great learning experience. Important aspects I found were:

Regarding my first question (Q1) on visualizing cohort data and on parallel coordinates in particular: Users took breaks between interaction steps to reflect on what they saw before continuing with incremental construction. Users did not express being overwhelmed even with a high amount of data. The think aloud method and results from the questionnaire show that the users interpreted the visualization very well. The *dialogues support* the interpretation and ensured that users spent more time with the visualization, i.e., motivated them looked back again at the visualization and think about it (aloud) before progressing further in the story.

Some subjects mentioned that it would be helpful to select individual polylines for a specific SHIP participant. So this is a technical optimization that I would definitely consider in future projects.

With question Q2 how to display time-oriented data with Parallel Coordinates I experienced some barriers with respect to the data set.

All participants expressed some level of confusion that the protagonist's liver values did not improve greatly over time as expected. Typical interpretations of this data visualization were, "The values have stayed the same or even gotten worse" or "I would have expected the values to have gotten better now". The interpretation was for this particular values correct as the Triglyceride value remained almost unchanged in the critical range, GGT value that was previously in the normal range ended up being at the upper limit of the critical range, with only the ASAT value having improved somewhat.



*Three time steps for the examination of the protagonist's liver values (SHIP)*

However, a typical time series includes several sampling points in time, with outliers being perceived in a weakened manner. So these data are not representative for this aim, especially if you take out just one study participant and then get three values over 15 years. This leads to much wrong scoring and confusion for a broad audience and it would therefore be recommended to exclude the visualization especially for an end of story.

Another idea for a temporal parallel coordinates plot could be visualizing the correlation between a risk factor, e.g., waist-to-height ratio, over the three time steps and health status of their liver. I would color the lines with respect to four groups: participants (1) recovering their fatty liver, (2) getting a fatty liver, (3) do not recover and (4) do not get fatty liver.

With respect to question Q3 how to design a cohort data stories, I summarized the most common aspects and lessons learned below:

First, users develop an emotional relationship to the protagonist, therefore I let the protagonist being present over most of the story after introducing them. For example, let Emma tell what the SHIP cohort study is and the doctor protagonist introduce and interpret/explain the diagrams.

Second, the users were interested in the AI generated characters and most expressed that they felt entertained. Because of technical barriers I could only produce short video sequences and autoplay with audio is not supported by

browser. However, most read short dialogs but would have appreciated longer text would have auditory support.

Third, users indicated that it was fun “taking own decisions”, “learn something”, the story was “well designed” and “the dialogues and interactivity were engaging”

Fourth, interaction elements could be improved, i.e., being more clear visible with their particular functionality.

Five, although users explore a layout in different ways, most users need a very clear structure in the layout and do not want to let their gaze wander over the layout too often. When I noticed this aspect, I discussed possible solutions with the users. Often the result was to arrange the layout elements even more according to the principle of western reading direction, e.g., to put the introduction dialog to a diagram on the left side, to leave the diagram in the middle, and to have the explanation dialog appear on the right side.

## Results from Descriptive Statistics

### (I) Memorability

As I said, participants did not know they would be asked about what content they remember from the story. 18 points was the maximum total score. With a mean  $M = 13.3$  and a standard deviation  $SD = 2.71$ , the result from the memorability test was with **73.8 %** satisfying ( $M = 13.3$ ,  $SD = 2.71$ ).

### (II) Comprehension

From questionnaire comprehension tasks a maximum total score of 11 points could be achieved, resulted in  $M = 7.7$  ( $SD = 2.31$ ), which is **70 %**.

### (III) User Engagement

Across all subcategories, there is an average Engagement with the communication tool of  $M = 6.068$  ( $SD = .65$ ), which is according to a high satisfaction (**75.85 %**).

## Results from Inferential Statistics

The calculation (ANOVA [Miller 1997]) for the factors surveyed shows that memorability, comprehension and user engagement are neither dependent on sex, age, visual impairment, professional knowledge in medicine or data visualization (not significant n.s.).

### (I) Memorability

sex ( $F_{(2,7)} = .184$ , n.s.), age ( $F_{(1,8)} = 1.845$ , n.s.), vis\_impair ( $F_{(2,7)} = .80$ , n.s.), med ( $F_{(2,7)} = .535$ , n.s.), datavis ( $F_{(2,7)} = .808$ , n.s.) and para\_coord ( $F_{(2,7)} = 1.32$ , n.s.).

### (II) Comprehension

sex ( $F_{(2,7)} = .067$ , n.s.), age ( $F_{(1,8)} = 1.597$ , n.s.), vis\_impair ( $F_{(2,7)} = .709$ , n.s.), med ( $F_{(2,7)} = .255$ , n.s.), datavis ( $F_{(2,7)} = .723$ , n.s.) and para\_coord ( $F_{(2,7)} = 2.793$ , n.s.).

### (III) User Engagement

sex ( $F_{(2,7)} = 2.134$ , n.s.), age ( $F_{(1,8)} = .059$ , n.s.), vis\_impair ( $F_{(2,7)} = .96$ , n.s.), med ( $F_{(2,7)} = .699$ , n.s.), datavis ( $F_{(2,7)} = 1.692$ , n.s.) and para\_coord ( $F_{(2,7)} = .141$ , n.s.).

## Correlation with Demographic Properties

No correlations between the three main categories Memorability, Comprehension and Engagement can be revealed.

This applies to Engagement\*Memorability ( $r = .243$ , n.s.) and Engagement \*Comprehension ( $r = .010$ , n.s.) as well Memorability\*Comprehension ( $r = -.002$ , n.s.).

The sample was small. However, with a larger sample this result could be interpreted as that the communication tool works well for a broad audience because it makes no difference in terms of gender, age, educational level, etc.

Accordingly, there is no covariance between the three main categories.

Subcategorically (enjoyment) there is only a significant positive correlation for FocusedAttention\*Memorability ( $r = .663$ ,  $p = .037$ ). I could conclude from this result that improving the story elements that increase focused attention also will improve memorability.

## 8 Conclusion & Future Work

As the results were satisfying, this story concept and prototype can be considered as starting point for developing a *template*, as the structure and data visualization techniques can generally be adapted to communicate other risk factors for specific health problems. For this purpose, new appropriate protagonists and alternatives to design the story can be implemented in the general structure.

This story concept can also be enriched. For example, some participants asked for more *personalization*, i.e., they would be interested to input their own data in a form and displaying their individual risk factors. “Where do I stand compared to others?” is a question that interests many users. I could imagine a highlighted polyline for the *user* emphasized in the foreground and on the background of *cohort data*.

With respect to the character design and gamification elements, some participants of my study felt remind of a computer game. I could imagine to integrate more advanced techniques of a point-and-click adventure game design into the concept.

To learn more about the difficulties and requirements of a broad audience, a *longitudinal study* would be very valuable to gain more insight into the perception of Parallel Coordinates by a heterogeneous target group. The number of subjects should be approximately equally distributed over gender, age and educational level to avoid bias.

As I have mentioned that *categorical data* cannot be depicted well with a classical Parallel Coordinate. For categorical data, e.g. physical activity with just two or four values, there must be an adapted concept, I could imagine combining parallel coordinates with parallel sets, and to dividing the axis into different sections analogous to the portion, with an icon assigned to each unique category value, for example high physical activity and low physical activity.

In the future, such a communication tool could be part of the *self-study*, e.g., in rehabilitation clinics or could be used in cooperation with health insurance companies or other health associations as a *preventive measure*.





## I. Bibliography

- Amini, Fereshteh, et al. “Hooked on data videos: assessing the effect of animation and pictographs on viewer engagement.” *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. 2018.
- Ashwell, Margaret, Pippa Gunn, and Sigrid Gibson. “Waist-to-height ratio is a better screening tool than waist circumference and BMI for adult cardiometabolic risk factors: systematic review and meta-analysis.” *Obesity reviews* 13.3 (2012): 275–286.
- Bachen, Christine M., et al. “How do presence, flow, and character identification affect players’ empathy and interest in learning from a serious computer game?.” *Computers in Human Behavior* 64 (2016): 77–87.
- Barlow, N., and Liz J. Stuart. “Animator: A tool for the animation of parallel coordinates.” *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004*. IEEE, 2004.
- Bateman, Scott, et al. “Useful junk? The effects of visual embellishment on comprehension and memorability of charts.” *Proceedings of the SIGCHI conference on human factors in computing systems*. 2010.
- Blaas, Jorik, Charl Botha, and Frits Post. “Extensions of Parallel Coordinates for interactive exploration of large multi-timepoint data sets.” *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008): 1436–1451.
- Borkin, Michelle A., et al. “What makes a visualization memorable?.” *IEEE transactions on visualization and computer graphics* 19.12 (2013): 2306–2315.
- Böttinger, Michael, Kostis, H-N., Velez-Rojas, Maria, Rheingans, Penny and Anders Ynnerman. “Reflections on visualization for broad audiences.” In *Foundations of Data Visualization*, pages 297–305. Springer, 2020.
- Burgers, Christian, et al. “How feedback boosts motivation and play in a brain-training game.” *Computers in Human Behavior* 48 (2015): 94–103.
- Campbell, Joseph. “The hero with a thousand faces.” Vol. 17. New World Library, 2008.
- Hazlett-Stevens H, Craske MG (2002) Brief Cognitive-Behavioral Therapy: Definition and Scientific Foundations. In *Handbook of Brief Cognitive Behaviour Therapy* (Eds F. W. Bond and W. Dryden), John Wiley & Sons Ltd, Chichester, UK.

- Cerioti, Ferruccio, et al. “Common reference intervals for aspartate aminotransferase (AST), alanine aminotransferase (ALT) and  $\gamma$ -glutamyl transferase (GGT) in serum: results from an IFCC multicenter study.” *Clinical chemistry and laboratory medicine* 48.11 (2010): 1593–1601.
- Chan, Winnie Wing-Yi. “A survey on multivariate data visualization.” *Department of Computer Science and Engineering. Hong Kong University of Science and Technology* 8.6 (2006): 1–29.
- Dykes, Brent. “Effective data storytelling: how to drive change with data, narrative and visuals.” *John Wiley & Sons*, 2019.
- Edsall, Robert M. “The parallel coordinate plot in action: design and use for geographic visualization.” *Computational Statistics & Data Analysis* 43.4 (2003): 605–619.
- Engelbrecht, Paula C., et al. “It is not worth learning if it is not remembered: designing e-learning to increase memory.” (2007).
- Górska, Renata. “Spatial imagination — an overview of the longitudinal research at Cracow University of Technology.” *Journal for Geometry and Graphics* 9.2 (2005): 201–208.
- Graham, Martin, and Jessie Kennedy. “Using curves to enhance parallel coordinate visualisations.” *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003*. IEEE, 2003.
- Gruendl, Henning, et al. “Time-Series Plots Integrated in Parallel-Coordinates Displays.” *Computer Graphics Forum*. Vol. 35. No. 3. 2016.
- Hefner, Dorothee, Christoph Klimmt, and Peter Vorderer. “Identification with the player character as determinant of video game enjoyment.” *Entertainment Computing-ICEC 2007: 6th International Conference, Shanghai, China, September 15–17, 2007. Proceedings*. Springer Berlin Heidelberg, 2007.
- Heinrich, Julian, and Daniel Weiskopf. “State of the Art of Parallel Coordinates.” *Eurographics (State of the Art Reports)* (2013): 95–116.
- Herpertz, U. “Adipositas-Diagnostik in der Lymphologie.” *LymphForsch* 13.2 (2009): 90–93.
- Hornsey, Matthew J., Josep Lobera, and Celia Díaz-Catalán. “Vaccine hesitancy is strongly associated with distrust of conventional medicine, and only weakly associated with trust in alternative medicine.” *Social Science & Medicine* 255 (2020): 113019.

- Inselberg, Alfred. “Parallel coordinates: intelligent Multivariate visualization.” *Intelligent Computer Graphics 2009* (2009): 123–141.
- Inselberg, Alfred, and Bernard Dimsdale. “Parallel coordinates: a tool for visualizing multi-dimensional geometry.” Proceedings of the first IEEE conference on visualization: visualization90. IEEE, 1990.
- Ittermann, Till, et al. “Inverse association between serum free thyroxine levels and hepatic steatosis: results from the Study of Health in Pomerania.” *Thyroid* 22.6 (2012): 568–574.
- Jiang, George J., and Tong Yao. “Stock price jumps and cross-sectional return predictability.” *Journal of Financial and Quantitative Analysis* 48.5 (2013): 1519–1544.
- Johansson, Jimmy, and Camilla Forsell. “Evaluation of parallel coordinates: Overview, categorization and guidelines for future research.” *IEEE transactions on visualization and computer graphics* 22.1 (2016): 579–588.
- Johansson, Jimmy, Patric Ljung, and Matthew Cooper. “Depth cues and density in temporal parallel coordinates.” *EuroVis*. Vol. 7. 2007.
- Keim, Daniel A., and H-P. Kriegel. “Visualization techniques for mining large databases: A comparison.” *IEEE Transactions on knowledge and data engineering* 8.6 (1996): 923–938.
- Kleinau, Anna, et al. “Is there a Tornado in Alex’s Blood Flow? A Case Study for Narrative Medical Visualization.” *Eurographics*, 2022.
- Klemm, Paul, et al. “Interactive visual analysis of lumbar back pain.” *Proc. of the 6th International Conference on Information Visualization Theory and Applications*. 2015.
- Kosara, Robert. “Presentation-oriented visualization techniques.” *IEEE computer graphics and applications* 36.1 (2016): 80–85.
- Kosara, Robert, and Jock Mackinlay. “Storytelling: The next step for visualization.” *Computer* 46.5 (2013): 44–50.
- Landay, James A., and Brad A. Myers. “Sketching storyboards to illustrate interface behaviors.” *Conference Companion on Human Factors in Computing Systems*. 1996.
- Lansdale, M. W. “On the memorability of icons in an information retrieval task.” *Behaviour & Information Technology* 7.2 (1988): 131–151.

Mar, Raymond A., et al. “Memory and comprehension of narrative versus expository texts: A meta-analysis.” *Psychonomic Bulletin & Review* 28 (2021): 732–749.

Meuschke, Monique, et al. “Towards narrative medical visualization.” *arXiv preprint arXiv:2108.05462* (2021).

Mittenentzwei, Sarah, et al. “Do Disease Stories need a Hero? Effects of Human Protagonists on a Narrative Visualization about Cerebral Small Vessel Disease Journal Article.” *Computer Graphics Forum*, 42 (3), 2023.

Miller Jr, Rupert G. *Beyond ANOVA: basics of applied statistics*. CRC press, 1997.

Mittenentzwei, Sarah, et al. “Narrative Visualization to Communicate Neurological Diseases.” *arXiv preprint arXiv:2212.10121* (2022).

Norboy o’g’li, Achilov Nurbek. “Pedagogical and psychological fundamentals of formation of space imagination and creative ability in students.” *European Journal of Research and Reflection in Educational Sciences* 8.4 (2020): 38–40.

O’Brien, Heather L., and Elaine G. Toms. “What is user engagement? A conceptual framework for defining user engagement with technology.” *Journal of the American society for Information Science and Technology* 59.6 (2008): 938–955.

Oddy, Wendy H., et al. “The Western dietary pattern is prospectively associated with nonalcoholic fatty liver disease in adolescence.” *Official journal of the American College of Gastroenterology/ ACG* 108.5 (2013): 778–785.

O’Donnell, Michael P. “Definition of health promotion 2.0: embracing passion, enhancing motivation, recognizing dynamic balance, and creating opportunities.” *American journal of health promotion* 24.1 (2009): iv–iv.

Oliveros, Estefania, et al. “The concept of normal weight obesity.” *Progress in cardiovascular diseases* 56.4 (2014): 426–433.

Pan, Bing, et al. “The determinants of web page viewing behavior: an eye-tracking study.” *Proceedings of the 2004 symposium on Eye tracking research & applications*. 2004.

Pawlow, Iwan Petrowitsch. “New researches on conditioned reflexes.” *Science* 58.1506 (1923): 359–361.

- Peña, Alyssa, Eric Ragan, and Lane Harrison. “Memorability of Enhanced Informational Graphics: the effects of design relevance and chart type on recall.” *Interdisciplinary Journal of Signage and Wayfinding* 4.1 (2020).
- Rasheed, Yawar, et al. “A model-driven approach for creating storyboards of web based user interfaces.” *Proceedings of the 7th International Conference on Computer and Communications Management*. 2019.
- Roland, Martin et al. (2007). Das Rückenbuch.
- Rombach, Robin, et al. “High-resolution image synthesis with latent diffusion models.” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
- Rosow, Irving. “What is a cohort and why?.” *Human Development* 21.2 (1978): 65–75.
- Schär, A., V. Messerli-Rohrbach, and P. Schubarth. “Conventional or complementary medicine: what criteria for choosing do patients use?.” *Schweizerische Medizinische Wochenschrift. Supplementum* 62 (1994): 18–27.
- Schorb, Friedrich. Body-Mass-Index. In: Anja Heermann et al. (Eds.), *Fat Studies: Ein Glossar (67–70)*. Bielefeld: transcript Verlag, 2022.
- Segel, Edward, and Jeffrey Heer. “Narrative visualization: Telling stories with data.” *IEEE transactions on visualization and computer graphics* 16.6 (2010): 1139–1148.
- Siirtola, Harri, et al. “Visual perception of parallel coordinate visualizations.” *2009 13th International Conference Information Visualisation*. IEEE, 2009.
- Skinner, Burrhus F. “Operant conditioning.” *The encyclopedia of education* 7 (1971): 29–33.
- So, Wonyoung, et al. “Humane visual ai: Telling the stories behind a medical condition.” *IEEE Transactions on Visualization and Computer Graphics* 27.2 (2020): 678–688.
- Theron, Roberto. “Visual analytics of paleoceanographic conditions.” *IEEE VAST*. 2006.
- Tominski, Christian, James Abello, and Heidrun Schumann. “Axes-based visualizations with radial layouts.” *Proceedings of the 2004 ACM symposium on Applied computing*. 2004.
- Tufte, Edward R. “The visual display of quantitative information.” (2001).

- Völzke, Henry, et al. "Cohort profile: the study of health in Pomerania." *International journal of epidemiology* 40.2 (2011): 294–307.
- Wang, Lujin, et al. "Color design for illustrative visualization." *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008): 1739–1754.
- Weitz, Darío. "Parallel Coordinates plots: Why & how, storytelling with parallels." *Towards Data Science post*: <https://towardsdatascience.com/parallel-coordinates-plots-6fcfa066dcb3> (2020).
- Wegenkittl, Rainer, Helwig Löffelmann, and Eduard Groller. "Visualizing the behaviour of higher dimensional dynamical systems." *Proceedings. Visualization'97 (Cat. No. 97CB36155)*. IEEE, 1997.
- Wegman, Edward J. "Hyperdimensional data analysis using parallel coordinates." *Journal of the American Statistical Association* 85.411 (1990): 664-675.
- Wing, Rena R., et al. "Behavioral science research in diabetes: lifestyle changes related to obesity, eating behavior, and physical activity." *Diabetes care* 24.1 (2001): 117–123.
- Withers, S. D. "Longitudinal methods (Cohort Analysis, Life Tables)." (2009): 285-292.
- Wong, Pak Chung, and R. Daniel Bergeron. "30 years of Multivariate multivariate visualization." *Scientific Visualization* 2 (1994): 3–33.
- Yang, Leni, et al. "A design space for applying the freytag's pyramid structure to data stories." *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2021): 922-932.
- Yasutake, K., Kohjima, M., Kotoh, K., Nakashima, M., Nakamuta, M., & Enjoji, M. (2014). Dietary habits and behaviors associated with nonalcoholic fatty liver disease. *World Journal of Gastroenterology: WJG*, 20(7), 1756.
- Zhang, Ying-xiu, et al. "Profiles of body mass index and the nutritional status among children and adolescents categorized by waist-to-height ratio cut-offs." *International Journal of Cardiology* 223 (2016): 529–533.

Figures:

Stefaner, Moritz "Visual Narratives: What's the story?" p. 23. 2012. url: <https://files.speakerdeck.com/presentations/508112265b45fd000202391f/virtual%20identity%20-%20Kaminabend%20-%20Moritz%20Stefaner%20-%20whats%20the%20story.pdf>

## II. Appendix

Table 4.2: Selection of Variables from SHIP-0 to SHIP-2

Variable (natural language)	Label	Datatype	Values of the Protagonist Subject no.: 2013160005002	Description and Measuring unit
<b>General</b>				
Gender/sex	<i>female_s0</i>	boolean	1	0 = male 1 = female
Age	<i>age_ship_s0</i>	ordinal	65	years
	<i>age_ship_s1</i>		70	
	<i>age_ship_s2</i>		77	
Date of data acquisition	<i>exdate_ship_s0</i>	ordinal	25.06.1998	Day of diagnosing fatty liver
Year of Birth (derived variable)	<i>year_birth</i>	ordinal	1933	[year ← exdate_ship_s0] – [age_ship_s0]
<b>Family</b>				
Marital status	<i>marit_s0</i>	categorical	2 = married, living separately from partner	↓
1 = married, live with partner, 2 = married, living separately from partner, 3 = single, never been married, 4 = divorced, 5 = widowed, 8 = do not know, 9 = refusal to answer				
Partnership	<i>partner_s0</i>	categorical	1 = living with a partner	↓
1 = yes, 2 = no, 8 = do not know, 9 = refusal to answer				
Birth	<i>birth_w_s0</i>	boolean	One or more children	at least one birth 0 = No 1 = Yes
<b>Education</b>				
Educational level	<i>edlevel_s0</i>	categorical	6 = Abitur	↓
1 = Hauptschule, 2 = Volksschule, 3 = derzeit Schüler/kein Abschluss, 4 = Mittlere Reife, 5 = Realschule, 6 = Abitur, 7 = Fachabitur, 8 = Fachhochschulreife				
<b>Occupation</b>				
Occupational group	<i>hf_kldb_75_s0</i>	categorical	864 = kindergarten teacher, childcare worker	KldB 1975 (Classification of occupations)
<b>Somatic Characteristics</b>				
Height	<i>som_groe_s2</i>	numerical	161	cm
Weight	<i>som_gew_s2</i>	numerical	60.6 (ideal weight)	kg

(Body weight for s0 and s1 are derived from bmi_s0 and bmi_s1)	<i>derived (s1)</i>	numerical	62.4	kg
	<i>derived (s0)</i>	numerical	69.8	kg
Diagnosis				
Fatty Liver (lat. Steatosis hepatis)	<i>stea_s0</i>	boolean	US pos.	US pos. = Fatty Liver, US neg. = No Fatty Liver, Diagnosed with ultrasound imaging
	<i>stea_s2</i>		US neg.	
Risk Factors for Fatty Liver				
BMI – Body Mass Index	<i>som_bmi_s0</i>	numerical	27.3 (overweight)	< 18.5 underweight 18.5–24.9 normal weight 25.0–29.9 overweight ≥ 30 obesity
	<i>som_bmi_s1</i>		24.4 (normal weight)	
	<i>som_bmi_s2</i>		23.3 (normal weight)	
Waist circumference	<i>som_tail_s0</i>	numerical	88.4	cm
	<i>som_tail_s1</i>		82.5	
	<i>som_tail_s2</i>		78.5	
Waist-to-Height Ratio ( <i>derived variable</i> )	<i>whtr_s0</i>	numerical	0.55	Waist-height Ratio ( <i>whtr</i> ) $= \frac{som\_tail\_s0}{som\_groe\_s2}$  0.4 < healthy whtr < 0.5
	<i>whtr_s1</i>		0.51	
	<i>whtr_s2</i>		0.49	
Balance of Nutrition	<i>ffs_s0</i>	ordinal	21 of 30	ffs = Food Frequency Score ↓
Dietary pattern	<i>ffs_pattern_s0</i>	ordinal (1, 2, 3)	3	Dietary pattern 1 = unfavourable 2 = intermediary 3 = optimal
Diabetes Mellitus Type 2	<i>diabetes_s0</i> <i>diabetes_s1</i> <i>diabetes_s2</i> <i>diab_type_s0</i>	boolean	0	Diabetes (known or diagnosed based on SHIP data) 0 = No 1 = Yes
	<i>diab_type_s0</i>	categorical	0	Diabetes Type 0 = No diabetes 1 = Diabetes type 1 2 = Diabetes type 2
Alcohol Problem	<i>alcprobl_s0</i>	boolean	0	0 = No alcohol problem 1 = Alcohol problem
Alcohol Consumption	<i>alkligt_s0</i>	numerical	3.36	Alcohol intake g/day (last week) [Alte 2004] if <i>alkligt</i> = 0 then abstinent
	<i>alkligt_s1</i>	numerical	2.28	
	<i>alkligt_s2</i>	numerical	0 = abstinent	
Smoking	<i>smoking_s0</i>	categorical	1	0 = Never Smoker



	<i>smoking_s1</i>			1 = Ex-Smoker 2 = Current Smoker
	<i>smoking_s2</i>			
	<i>onsetsmok_s0</i>	ordinal	17	<i>years</i> , Age of smoking onset (first smoking)
<b>Mental Health</b>				
Anxiety	<i>anxiety_s0</i>	boolean	1	0 = No 1 = Yes (See annex for definition)
Depression	<i>depre_s0</i>	boolean	0	0 = No 1 = Yes (See annex for definition)
<b>Healthcare</b> (indicator for Self-care)				
Visits to the doctor (last 4 weeks)	Visits to the doctor (last 4 weeks)	ordinal	1	Visits to the doctor (last 4 weeks)
Visits to the doctor (last 12 month)	Visits to the doctor (last 12 month)	ordinal	13	Visits to the doctor (last 12 month)

Table 5.2: Liver Function Values from SHIP-0 to SHIP-2

Variable (natural language)	Label	Datatype	Value of the Subject (Protagonist)	Reference Range
<i>Liver Function Values</i>				
Triglyceride (TG)	tg_s_s0	numerical	2.63	< 1.9 mmol/l
	tg_s_s1		2.67	
	tg_s_s2		2.57	
Gamma-Glutamyl-Transferase (GGT)	ggt_s_s0	numerical	0.44	< 0.96 $\mu$ katal/l
	ggt_s_s1		0.52	
	ggt_s_s2		0.93	
Aspartate aminotransferase (ASAT) Alternative name: glutamic oxaloacetic transaminase (GOT)	asat_s_s0	numerical	0.27	< 0.59 $\mu$ katal/l
	asat_s_s1		0.21	
	asat_s_s2		0.16	
Alanine transaminase (ALAT/ALT/GPT)	alat_s_s0	numerical	0.39	< 0.77 $\mu$ katal/l
	alat_s_s1		–	
	alat_s_s2		–	
LDL Cholesterol	ldl_s_s0	numerical	4.02	< 3.34 mmol/l
	ldl_s_s1		2.59	
	ldl_s_s2		4.16	
HDL Cholesterol	hdl_s_s0	numerical	1.43	> 1.03 mmol/l
	hdl_s_s1		0.8	
	hdl_s_s2		1.47	

# Fragebogen

## Persönliche Angaben

☐ weiblich      ☐ männlich      ☐ divers      ☐ keine Angabe

Alter:

☐ 18–25      ☐ 26–35      ☐ 36–45      ☐ 46–55      ☐ 56–65      ☐ > 65

Höchster Bildungsabschluss:

☐ Schüler\*in      ☐ kein Abschluss      ☐ Abitur/Hochschulreife      ☐ Berufsausbildung  
☐ Fachhochschulabschluss      ☐ Hochschulabschluss

Defizite im Sehen:

☐ Keine Sehschwäche      ☐ Farbsehschwäche  
☐ Kurzsichtigkeit      ☐ Weitsichtigkeit

Ich beschäftige mich beruflich oder im Studium mit medizinischen Themen.

☐ Ja      ☐ ein bisschen/manchmal      ☐ Nein

Ich beschäftige mich beruflich oder im Studium mit Datenvisualisierungen.

☐ Ja      ☐ ein bisschen/manchmal      ☐ Nein

Ich kenne Parallele Koordinaten (Parallel Coordinates) als Visualisierungstyp.

☐ Ja      ☐ ein bisschen/manchmal      ☐ Nein

Bei mir wurde eine Fettleber diagnostiziert.

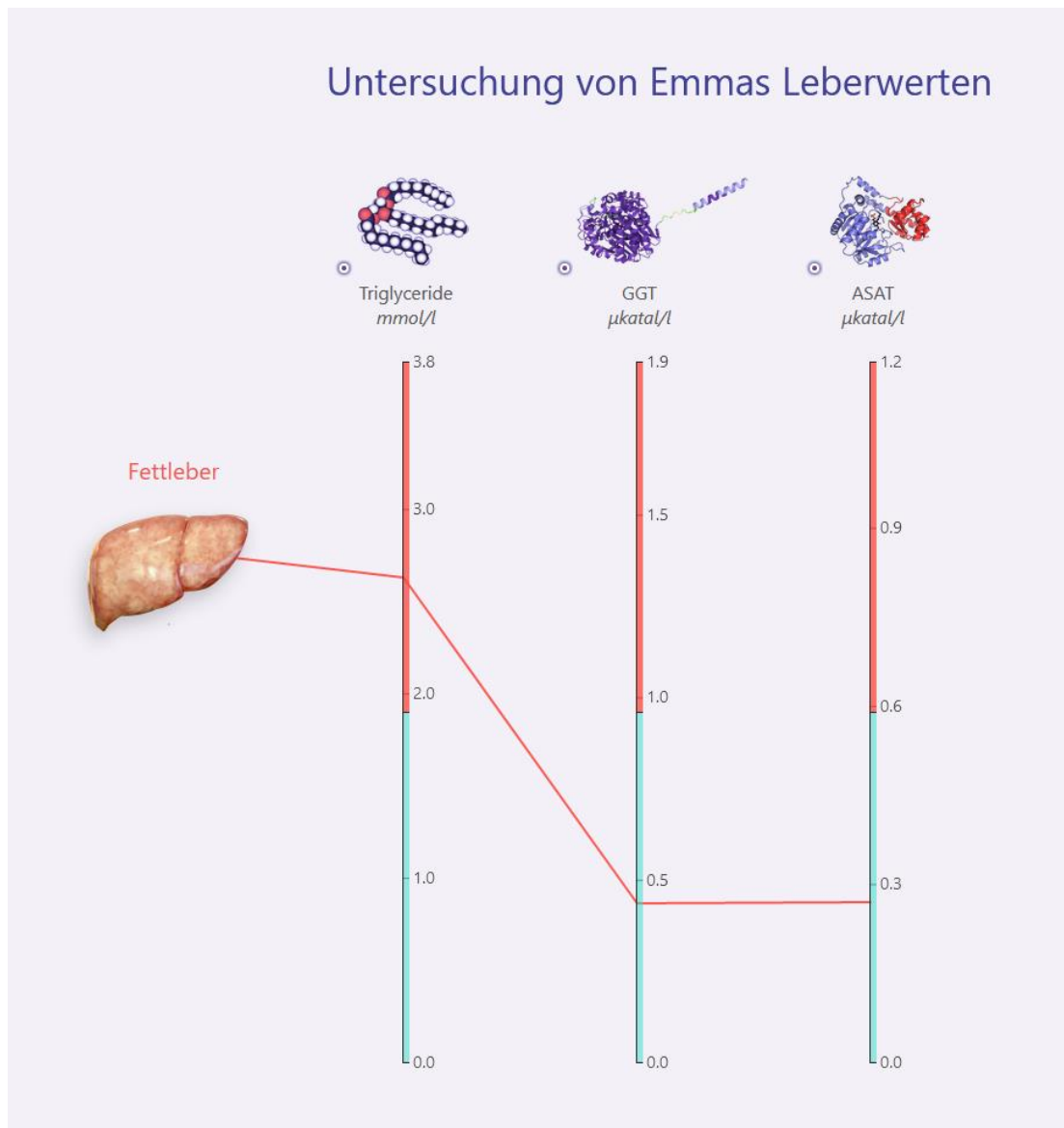
☐ Ja      ☐ Nein

## **(I) Inhaltliche Fragen zur Geschichte**

(Kreuzen Sie eine oder mehrere Antworten an! (Mehrfachauswahl))

1. Wie heißt die Protagonistin in der Geschichte?  
☐ Emma Wind      ☐ Erna Winter      ☐ Emma Winter
2. Welche medizinischen Untersuchungstechniken wurden in der Geschichte für die Diagnose einer Fettleber eingesetzt?  
☐ Ultraschall    ☐ Magnet-Resonanz-Tomographie    ☐ Wiegen    ☐ Bluttest
3. Was sind die Merkmale einer Fettleber?  
☐ Vergrößerung der Leber    ☐ helleres Erscheinungsbild im Ultraschallbild  
☐ 10% der Leberzellen enthalten Fett    ☐ 5% der Leberzellen enthalten Fett
4. Was sind vermeidbare Risikofaktoren für eine Fettleber?  
☐ Alkoholkonsum    ☐ Bauchfett    ☐ BMI    ☐ Alter
5. Welche Folgeerkrankungen einer Fettleber wurden in der Geschichte genannt?  
☐ Leberfibrose      ☐ Leberzirrhose      ☐ Leberkrebs
6. Welche Funktionen hat eine Leber?  
☐ Speicherfunktion      ☐ Entgiftung      ☐ Stoffwechsel      ☐ Immunabwehr
7. Was sind wichtige Verhaltensänderungen eines gesünderen Lebensstils in Bezug auf eine Leber?  
☐ Immunabwehr steigern    ☐ Umstellung der Diät    ☐ Geistige Aktivität steigern  
  
☐ Alkoholkonsums minimieren      ☐ Körperliche Aktivität steigern

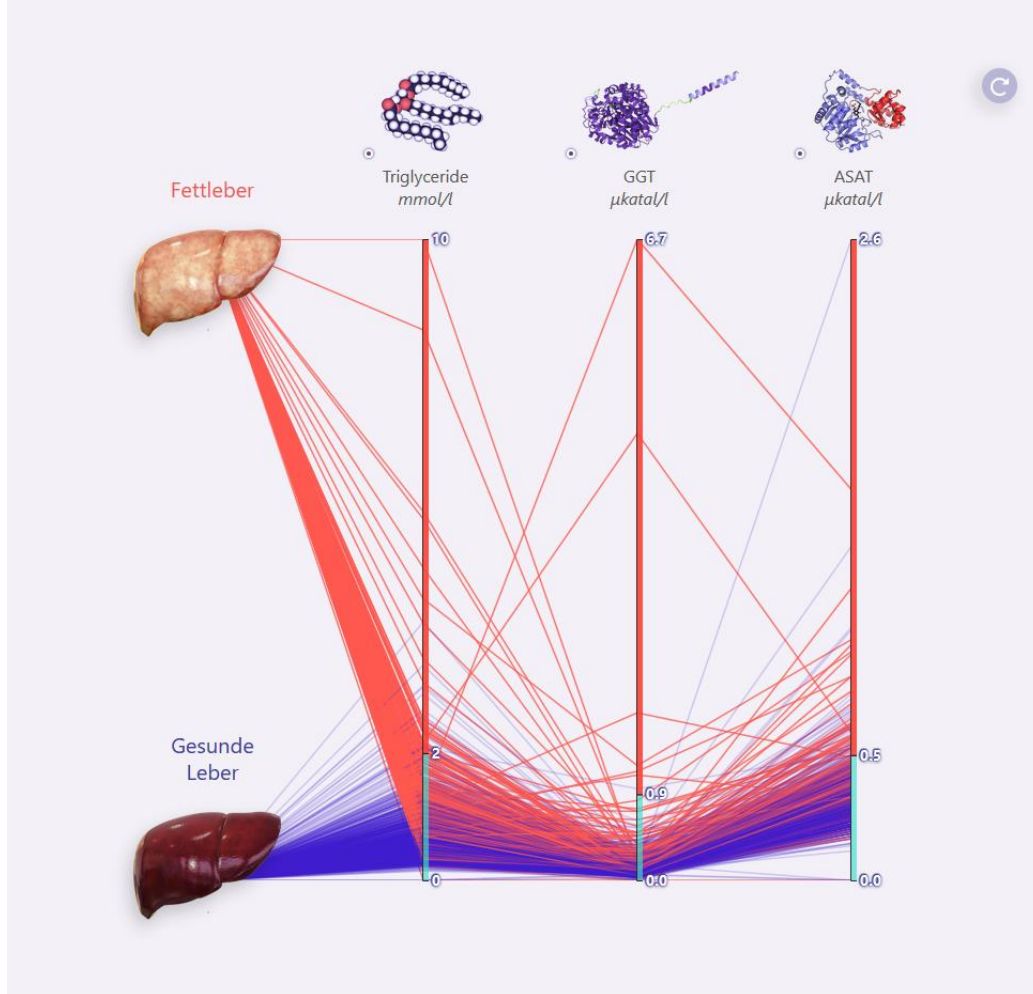
## (II) Fragen zu Parallelen Koordinaten (Multiple Choice)



Welche Leberwerte sind im Normbereich und im kritischen Bereich?

	Normbereich	Kritischer Bereich
Triglyceride	<input type="checkbox"/>	<input type="checkbox"/>
GGT	<input type="checkbox"/>	<input type="checkbox"/>
ASAT	<input type="checkbox"/>	<input type="checkbox"/>

## Leberwerte aus Kohortendaten



Wofür steht eine Linie?

- ☐ Leberwert      ☐ Studienteilnehmer

Was stellt eine Achse dar?

- ☐ Leberwert      ☐ Studienteilnehmer

Wofür steht eine rote Linie?  
(freie Antwort)

Wofür steht eine blaue Linie?  
(freie Antwort)

Welchen Stoff kann der Körper selbst  
(aus Zucker und Alkohol) herstellen?

Triglyceride	<input type="checkbox"/>
GGT	<input type="checkbox"/>
ASAT	<input type="checkbox"/>

Welcher Stoff kommt in der Nahrung vor?

Triglyceride	<input type="checkbox"/>
GGT	<input type="checkbox"/>
ASAT	<input type="checkbox"/>

Welche Leberwerte weisen auf eine  
Leberentzündung hin?

Triglyceride	<input type="checkbox"/>
GGT	<input type="checkbox"/>
ASAT	<input type="checkbox"/>

### (III) Fragen zum Erleben der Geschichte

(Kreuze eine Zahl an, die am ehesten/besten auf die Aussage zutrifft)

#### (A) Affective Involvement

Ich habe mich mit der Protagonistin emotional verbunden gefühlt.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

Ich hatte das Gefühl, aktiv an der Geschichte beteiligt zu sein.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

Der Anfang der Geschichte hat mich zum Weitermachen motiviert.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

#### (B) Enjoyment

Die Geschichte war unterhaltsam.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

Was hat Spaß gemacht? (freie Antwort)

Die Geschichte war zu lang

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

#### (C) Aesthetics

Die folgenden Elemente fand ich ästhetisch:

Bilder von den Protagonisten

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

Symbole / Icons

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

3D-Visualisierungen

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

(D) Focused attention

Ich habe mein Zeitgefühl verloren, als ich die Geschichte sah.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

Während ich die Geschichte sah, schweiften meine Gedanken ab.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

Es fiel mir schwer, mich auf die Geschichte zu konzentrieren.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

(E) Cognitive Involvement

Die Geschichte war fesselnd.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

Die Struktur der Geschichte hat mich verwirrt.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

Der Inhalt der Geschichte war leicht zu verstehen.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

Ich habe etwas Neues gelernt.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

(F) Interaktion

Die Interaktion mit den Datenvisualisierungen/3D-Darstellungen ist nützlich,  
um den Inhalt zu verstehen

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

Ich wusste sofort, wie ich mit den Visualisierungen interagieren sollte.

Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu

Ich hatte Spaß an der Interaktion.

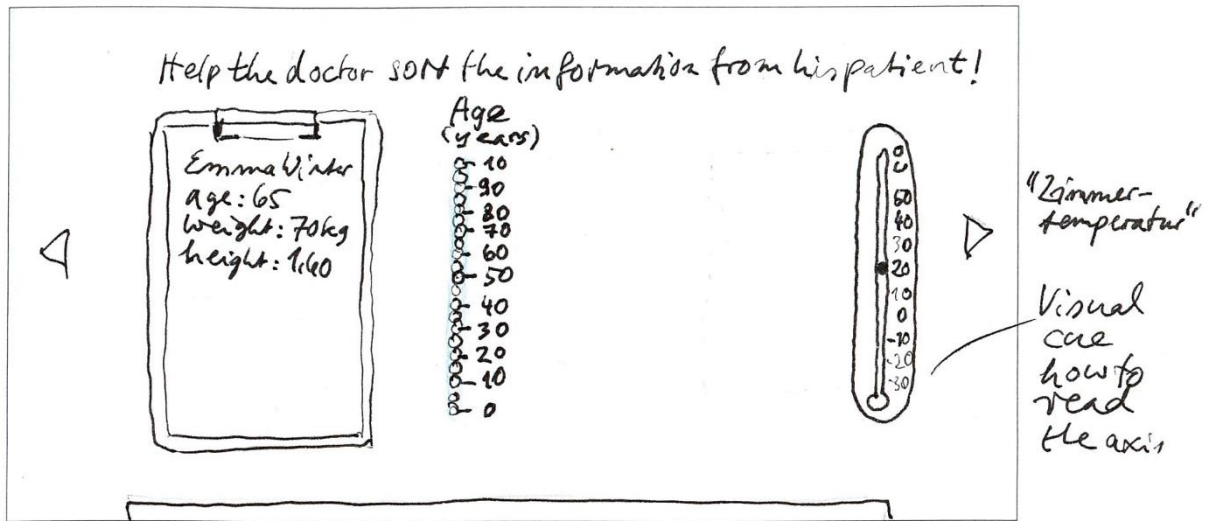
Trifft gar nicht zu    1   2   3   4   5   6   7   8    Trifft vollkommen zu





1

## Introductory Example to Parallel Coordinates



Protagonists:

Description:

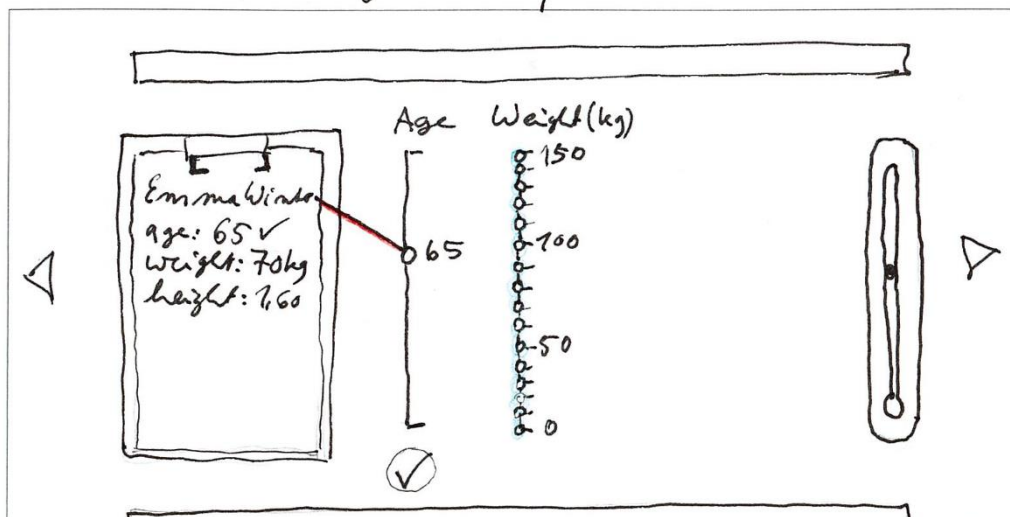
Text:

UI:

Introduction to Incremental construction at the same time, working a list through = motivation technique  
 using a thermometer as visual cue  
 (metaphor), UI: Buttons glowing effect

2

## Introductory Example to Parallel Coordinates



Protagonists:

Description:

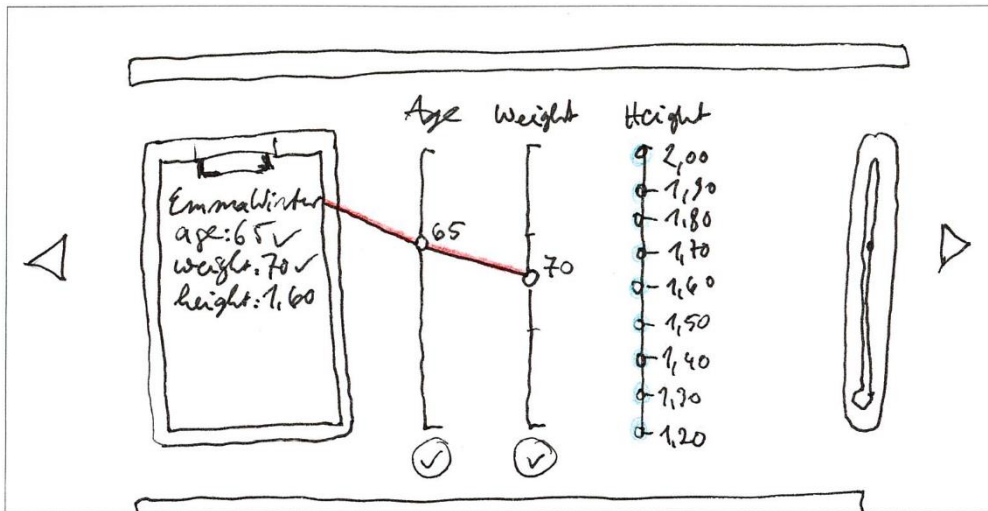
Text:

UI:

Working through a list is a motivation technique  
 Done things get a check mark

3

## Introductory Example to Parallel Coordinates



Protagonists:

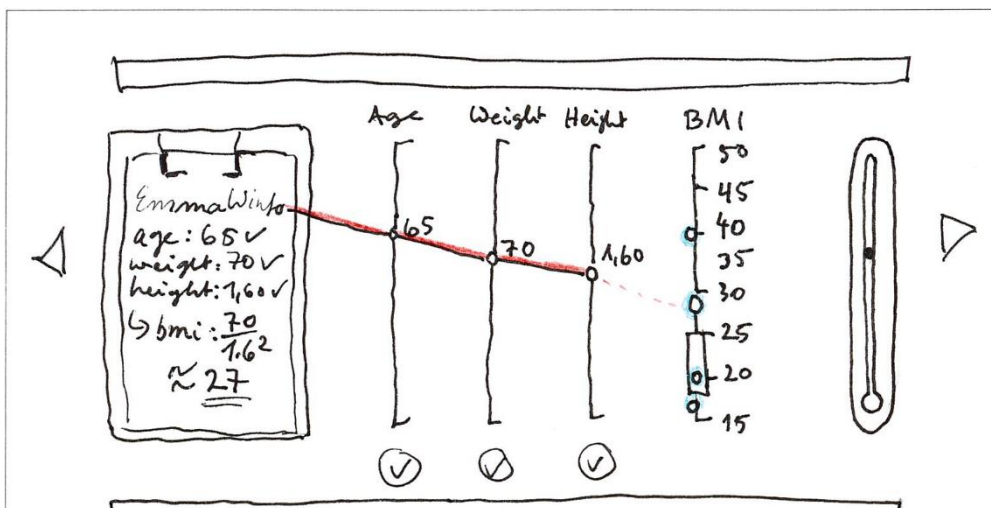
Description:

Text:

UI:

4

## Introductory Example to Parallel Coordinates



Protagonists:

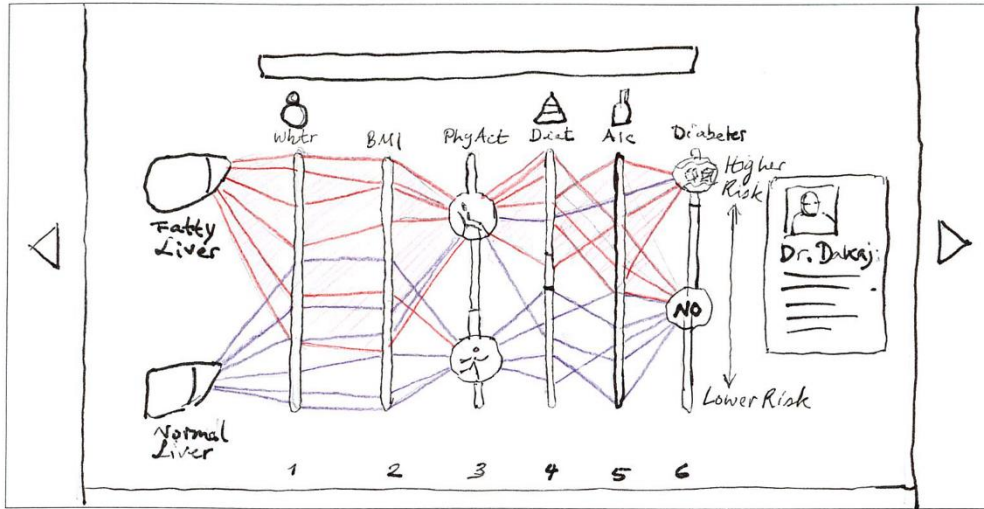
Description:

Text:

UI:



## Parallel Coordinates for Risk Factors (Time Step 1)



Protagonists: "Non-alcoholic fatty liver" risk factor

Description: Gradient<sup>2,5</sup> Sections<sup>2,4</sup> Binary<sup>3,6</sup> (Distribution is shown on the vertical bar/axis)  
the values on the top represent the 'risk factors'

Text: that is high waist-to-height ratio, high BMI,

UI: especially obesity, low physical activity, unbalanced diet, higher alcohol consumption and diabetes mellitus Type 2

Protagonists:

Description:

Text:

UI:





## **Selbstständigkeitserklärung**

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst und keine anderen Hilfsmittel als die angegebenen verwendet habe.

Insbesondere versichere ich, dass ich alle wörtlichen und sinngemäßen Übernahmen aus anderen Werken als solche kenntlich gemacht habe.

Magdeburg, der 9.6.2023

.....