Institute for Visualization and Interactive Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Masterarbeit

Visual Analytics of Geolocated Open Fitness Data of Children and Teens

Oliver Richard Ferch

Course of Study:	Software Engineering
Examiner:	Dr. Tanja Blascheck
Supervisor:	Dr. Tanja Blascheck, Dr. Claudia Niessner (KIT)
Commenced:	June 8, 2021
Completed:	December 8, 2021

Abstract

The Motorik Modul (MoMo) study conducted by the Karlsruhe Institute of Technology (KIT), collected a range of physical activity data of children and teenagers in Germany across multiple time waves to analyze underlying effects on their fitness and health over a large period of time. The physical activity data consists of BMI scores, accelerometer data which shows how active participants were during the day and various fitness tests, that were measured to determine the participant's fitness status. In this work the Visual Analytics of Sport Data (VASD) system is introduced, which visualizes different aspects of this data for a holistic analysis. With this system an analysis across one or multiple time waves is possible and the analysis is done on groups of participants in states of Germany or for one or multiple participants of a selected state. There were some longitudinal participants who participated in more than one of the MoMo time waves, which allows an analysis of their performance over time. Using the different data categories about Accelerometer, Fitness, BMI and Questionnaire with the various linked charts that visualize the data, for example pie chart and scarf plot, the user can analyze specific fitness and health trends that appear in the data. VASD offers an overview of the collected MoMo data and detailed analyses for selected participants. An evaluation of a small user study conducted for the system concludes that VASD improves the comprehensibility of the data set. In this thesis, three use cases are presented to demonstrate the possibilities of using the system.

In der vom Karlsruher Institut für Technologie (KIT) durchgeführten Motorik Modul (MoMo) Studie wurde eine Reihe von Daten über die körperliche Aktivität von Kindern und Jugendlichen in Deutschland über mehrere Zeitwellen hinweg gesammelt, um die zugrunde liegenden Auswirkungen auf ihre Fitness und Gesundheit über einen langen Zeitraum zu analysieren. Die Daten zur körperlichen Aktivität bestehen aus BMI Werten, Akzelerometerdaten, die zeigen, wie aktiv die Teilnehmer während des Tages waren und verschiedenen Fitnesstests, die zur Bestimmung des Fitnessstatus der Teilnehmer gemessen wurden. In dieser Arbeit wird das Visuelle Analyse von Sportdaten (VASD) System vorgestellt, das verschiedene Aspekte dieser Daten für eine ganzheitliche Analyse visualisiert. Mit diesem System ist eine Analyse über einen oder mehrere Zeitwellen möglich und die Analyse wird für Gruppen von Teilnehmern in den Bundesländern oder für einen oder mehrere Teilnehmer eines ausgewählten Bundeslandes durchgeführt. Es gab einige Längsschnitt Teilnehmer, die an mehr als einer der MoMo Zeitwellen teilgenommen haben, was eine Analyse ihrer Leistung über die Zeit ermöglicht. Mit Hilfe der verschiedenen Datenkategorien zu Akzelerometer, Fitness, BMI und Fragebogen mit den verschiedenen verknüpften Diagrammen, die die Daten visualisieren, zum Beispiel Tortendiagramm und Streudiagramm, kann der Benutzer spezifische Fitness– und Gesundheitstrends analysieren, die in den Daten auftauchen. VASD bietet einen Überblick über die gesammelten MoMo Daten und detaillierte Analysen für ausgewählte Teilnehmer. Eine Auswertung einer kleinen Nutzerstudie, die für das System durchgeführt wurde, kommt zu dem Schluss, dass VASD die Verständlichkeit des Datensatzes verbessert. In dieser Arbeit werden drei Anwendungsfälle vorgestellt, um die Möglichkeiten der Nutzung des Systems zu demonstrieren.

Contents

List of Figures

List of Tables

Acronyms

ACM Association for Computing Machinery. 26

AOIs Areas of Interest. 21

API Application Programming Interface. 29

BMI Body Mass Index. 13

BMI IOTF Body Mass Index according to International Obesity TaskForce. 39

BMI KH Body Mass Index according to Kromeyer-Hauschild. 39

Bpm Beats per minute. 24

ECG Electrocardiogram. 24

ID Identifier. 49

IEEE Institute of Electrical and Electronics Engineers. 25

KiGGS German Health Interview and Examination Survey for Children and Adolescents. 15

KIT Karlsruhe Institute of Technology. 13

MAT Microsoft Access Table Shortcut. 30

MoMo Motorik Modul. 13

PWC Physical Working Capacity. 33

SAV Statistical Package for the Social Sciences Statistics Data File Format Family. 30

SUS System Usability Scale. 52

TTT Type by Task Taxonomy. 17

VASD Visual Analytics of Sport Data. 13

VIS Institute for Visualization and Interactive Systems. 13

1 Introduction

The Motorik Modul (MoMo) study collected fitness and health data from children and teenagers in Germany to analyze their physical fitness and capabilities, with the different circumstances of each child, if it is doing sport in a club or any other outside activities. The researchers of Karlsruhe Institute of Technology (KIT) wanted to see if an active child had better fitness scores and abilities as a child which is less active in their free time. The data collected was then analyzed using manual methods, like formulas and filters to receive answers to hypotheses, for example if a child is active in a young state, do they stay active in their later life? Therefore, children who participated in multiple time waves are long time participants of this study and can be analyzed further.

After some time with the collected data the demand increased, to try to visualize all this data, even for outside users, which do not have deep insights into the type of collected data. The visualizations should give users a quick way to get a grasp on the data, for example to check for possible seen effects of the corona pandemic on the physical physique of the children, caused by the lockdowns. For this idea, the KIT cooperated with the Institute for Visualization and Interactive Systems (VIS) of the University of Stuttgart, to create a system visualizing the MoMo data for overview analysis and more detailed analysis. The system created and presented in this thesis is called Visual Analytics of Sport Data (VASD).

VASD is realized as a website to easily distribute it to users without the need of any work on their side. Four categories are available to choose from, which all target different parts of the MoMo data. These parts are accelerometer data, fitness tests, Body Mass Index (BMI) scores and questionnaire answers. Filtering of the data is supported through using several filters, for example age, gender or city size. Time waves T1–T6 can be selected for the categories Fitness, BMI and Questionnaire and a month range selection is available. For category Accelerometer, only time wave T3 is available and a date range selection, where single days can be selected. Choosing one of the categories also changes some linked visualizations, for example pie charts for each selected participant, to highlight information from multiple sides and give a more detailed look at it. VASD was evaluated with a small user study and this thesis presents three use cases that show the possibilities that the system offers for analyzing the data.

The following chapters in this thesis are used to describe the influencing factors and considerations on how the VASD system was created. Chapter 2 is about the Related Work of different visualizations used in the system, the fitness data processed and systems that visualize different types of data in a similar but distinct way. In Chapter 3 the concept and implementation of VASD are explained. An Evaluation of the system is done in Chapter 4 with a small user study. Chapter 5 describes use cases for which VASD can be used to gain new insights or a general overview on the MoMo data set. In Chapter 6 there is a conclusion on VASD with limitations of the system and future work which can be done to improve the system further.

2 Related Work

This section discusses related work that influenced the decision making for the implementation of VASD. The origin of the MoMo data is explained, the visualization theory is presented, the charts used in the system are described in detail, examples of visualization systems are shown, and a distinction is made between this work and the related work.

2.1 Fitness and Health Data in the MoMo Study

Doing sports to increase a person's health is a widely known fact in the public health literature [BMA+20]. For children, it is difficult to measure their level of physical activity because they are more likely to do their types of activity more randomly than adults. A very active child might use more energy than an adult who does not exercise regularly [BMA+20]. Questionnaires are used by KIT researchers in the MoMo study to get a basic line of knowledge about children's physical activity. With the additional use of accelerometers, they try to get more detailed information about the physical activity of children. In the MoMo questionnaire participants were asked about social environmental determinants [BMA+20], for example the infrastructure of their neighborhoods, to see if they were a deciding factor on how fit children were. Because a child with a lot of green space or public spaces to play could be fitter than a child with little to no green space or public spaces around them. Additionally, the MoMo project conducted a longitudinal study to track the change of physical activity and health across different time waves in the lives of participants, which took part in multiple waves of the study.

The MoMo project uses the „Studie zur Gesundheit von Kindern und Jugendlichen in Deutschland" – German Health Interview and Examination Survey for Children and Adolescents (KiGGS) from the Robert Koch Institute [Ins] as their baseline to get a part of their participants and add physical activity monitoring to increase the information width. MoMo selected 4,538 participants from the years 2003–2006 aged 4 to 17 years, from wave 2009–2012 5,106 participants aged 4 to 23 years and from wave 2014–2017 5,689 participants aged 4 to 30 years, where accelerometer data for the first time was tracked [BMA+20]. There is also a current wave from 2018–2020, which includes questionnaires for the corona time.

From 4,538 invited persons to wear accelerometers and collect accelerometer data, 1,974 agreed to participate and had their data downloaded from the devices. Eight days of accelerometer data was tracked, the first day of data was not used to dismiss confusion of participants wearing the device, which resulted in one whole week of useful data. Getting the non–wear times from participants was a challenge, because the protocol children had to fill out was often not accurate enough (children forgot to enter exact times or forgot to enter them completely). Nevertheless, to process this unfiltered accelerometer data and receive wear and non–wear times the Choi algorithm was used, for more details see Section 3.3.2.

Other measurements performed for MoMo were fitness tests to gain insights into the fitness and health status of the participants. Eleven tests were conducted and included measurements of „endurance, strength, fine and gross motor coordination, and flexibility" [WKB+21].

One of the findings in the MoMo Study was that physical activity was almost the same between 2003 and 2017, but time spent outside for children on their own decreased while time in sports clubs increased [WKB+21]. Physical fitness levels were better in 2009–2012 compared to 2003–2006, but did not change afterwards in comparison to 2015–2017 [WKB+21].

To simplify when talking about different waves of the MoMo study, this thesis uses the term time wave with the number of the wave. The short version for time wave 1 is T1. The time waves used in this thesis with their investigation periods are:

- T1 (2003–2006)
- T2 (2009–2012)
- T3 (2014–2017)
- T4 (2018–2020)
- T5 (2020–2021)
- T6 (2021–2021)

2.2 Visualization and Visual Analytics

With the use of visualization huge amounts of data can be analyzed. Selecting the right types of visualizations is an important step in meaningful data analysis. The visualizations must fit the data and the type of research an analyst wants to do, „to pursue questions, uncover patterns of interest, and identify (and potentially correct) errors" [HS12].

To show different visualizations, the underlying data set needs to be processed and mapped to the visual representation. Card et al. [CMS99] introduced a visualization pipeline, see Figure 2.1. Raw data gets processed and saved in files or data tables for later use, this data also can get filtered through interaction. This processed data is mapped to visualizations and lastly viewed by an analyst, who does various interactions with the system to explore the data.
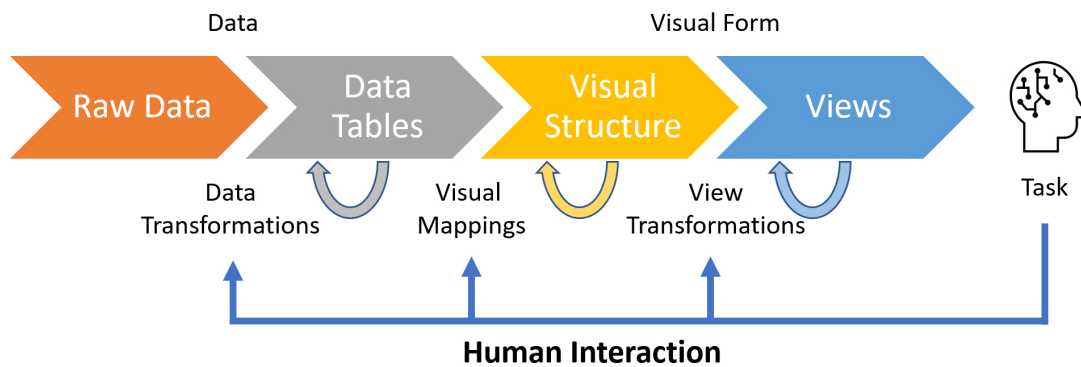
Figure 2.1: Steps of visualization processing, figure based on Card et al. [CMS99] visualization pipeline. A data set is processed into data tables which hold the relations of the data. This data is then mapped to the visual structures, that contain graphical properties and marks. Transforming this data to the views then gives graphical parameters, for example position and scaling [CMS99].

Shneiderman introduced the Visual Information–Seeking Mantra for designing graphical user interfaces: „Overview first, zoom and filter, then details–on–demand" [Shn03] (Figure 2.2). For information visualization, various data types appear which have different tasks that need to be handled, Shneiderman defines this as the Type by Task Taxonomy (TTT). The seven tasks defined for TTT are [Shn03]:

- **Overview**: An analyst gets an impression about the whole collection of data available.
- **Zoom**: An analyst can select single items of interest and focus on them.
- **Filtering**: Removes unwanted data and shows remaining data points to the analyst.
- **Details–on–Demand**: Show more detailed information for single items or a group to the analyst.
- **Relate**: An analyst can see relationships between linked items.
- **History**: An analyst's actions can be reverted or done again more refined.
- **Extract**: An analyst can extract information from selections.

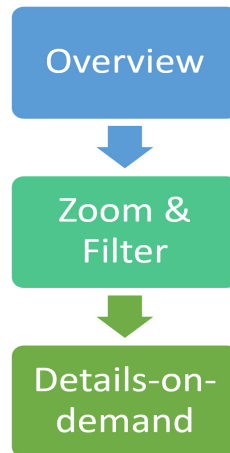**Figure 2.2:** A visualization first gives an overview of the data set, then filtering and selection of the data is done. Finally, more detailed information is presented for the selected part of the data set. Figure based on Shneiderman's Visual Information–Seeking Mantra [Shn03].

2.2.1 Pie and Donut Chart

This section explains the use of pie and donut charts, why they were used in VASD and shows related work for the use of pie and donut charts.

Pie Chart

Pie charts are an effective way to display relations between available data points and help in comparing them and finding the maximum value from all the divisions displayed. Each slice of the pie chart represents a data point and its value, in comparison to the whole pie chart, see an example in Figure 2.3.

In the publication of Saket et al. [SED18] the use of different visualization techniques was tested for various data analysis tasks. The five visualizations tested were Table, Line Chart, Bar Chart, Scatter plot and Pie Chart [SED18]. Pie charts were as good as bar chart and table „for Retrieve, Range, Order, Filter, Extremum, Derived and Cluster tasks, it is less accurate for Correlation, Anomalies and Distribution tasks" [SED18].

Values of the data set are hard to retrieve in pie charts and because of this, labels are added to it [SED18]. But the use of labels brings the downside of too much visual information displayed. The use of a legend for the explanation of the different slices helps reduce the visual clutter [SED18]. With the use of a tool–tip the exact values of the slices can be shown on selection.
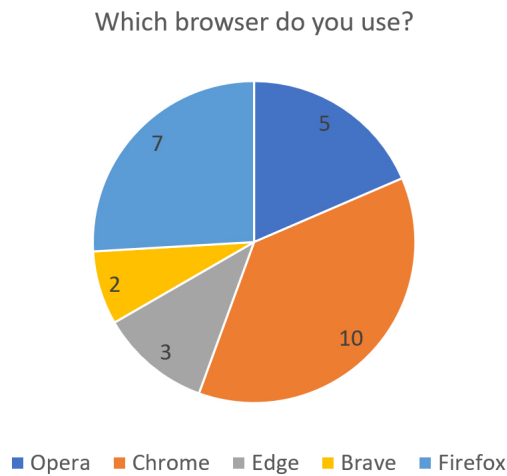
Which browser do you use?

Figure 2.3: Pie chart example of a survey about the internet browsers that participants use

The pie chart visualization was added to VASD because it offers fast relation comparisons between data points, which allows the system to give a first overview of the participants data. Bar charts would help with retrieving exact values, but the overview with pie charts is enough to see relative differences in all the data points.

In the publication of Overbye and Weber [OW00] pie charts were used on a geographic map to visualize the usage of transmission lines for a power system. With a dynamic size and color of the pie charts and a label that presents the percentage of loading on each line, an overview was possible on the map [OW00]. One drawback of this approach was that too many pie charts would clutter the map, making it harder to distinguish and find information.

Donut Chart

A donut chart is another form of visualization for a pie chart. The large amount of participants in MoMo made the use of pie charts on the geographic map more difficult. To deal with this amount of data a grouping of the data was needed and led to the use of donut charts. Pie charts are only displayed for single participants in an area or when participants are selected for more details on each participant.

In the paper by Blascheck et al. [BBB+18] two smartwatch studies were conducted comparing bar charts with donut and radial charts for a data comparison task, see Table 2.1. Participants of the first study ranked the different charts „based on their preference and their level of confidence in performing the task correctly" [BBB+18] and donut charts were just ahead of bar charts in both categories for all data sizes. In a follow up study the donut chart was again most preferred before the bar chart with one exception compared to the first study at seven data values, where the bar chart was more preferred than the donut chart [BBB+18]. Donut and bar charts performed similarly, with only small differences in speed tests for retrieving data from the charts.
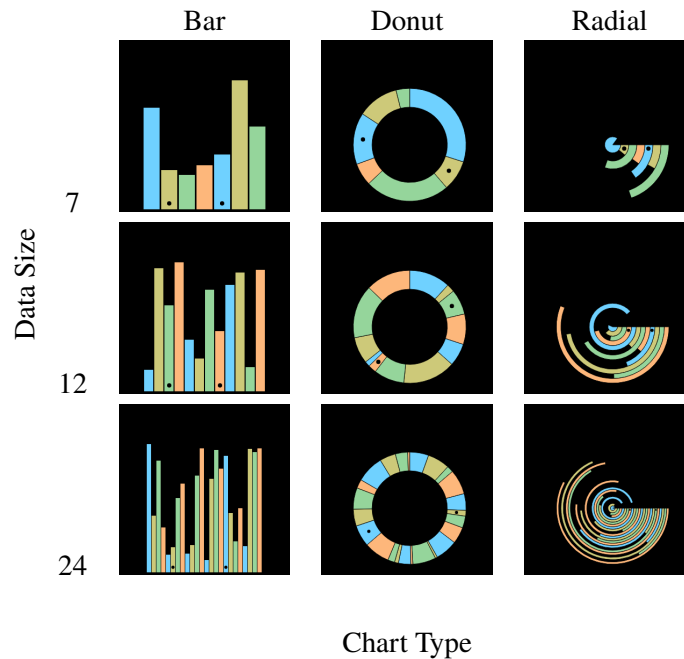
Table 2.1: Visualization examples used in the study of Blascheck et al. [BBB+18] with three different data sizes.

In a paper by Aravind et al. [ABI19] visualizations for fitness trackers were tested including donut charts that visualized the sleep duration. The big visualization of a donut chart was the preferred design compared to a bar chart, which was more preferred in a smaller visualization. A publication by Peterson et al. [PSBS19] used donut charts to compare participants gaze data with each other.

The results of these publications show that the use of donut charts is useful in giving an overview of the data set and is slightly more preferred compared to bar charts. Because the pie chart is used in VASD for a more detailed analysis of the data, the use of donut charts offers itself to show the groups of participants, as it keeps the visualization consistent.

2.2.2 Radar or Star Chart

Radar or star charts display data points on multiple axes with a different variable tied to each axis. Differences in data points can be compared with each other when using multiple radar charts, or in one single radar chart with multiple time waves displayed.

In VASD the radar chart is used, because various fitness tests are available in the MoMo data set as their own variables, which allows for comparisons across multiple time waves or with other participants.

In the paper by Munoz et al. [MBRC15] a radar chart is used to display multiple data points for three or more quantitative variables, which are projected on different axes [MBRC15], see Figure 2.4.
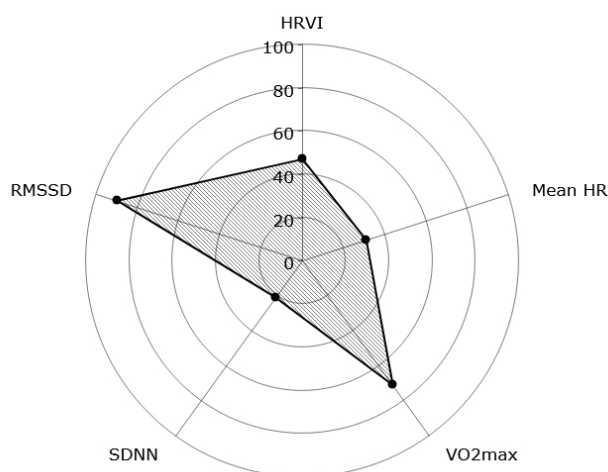
Figure 2.4: Radar chart which shows the different data points for all axes, figure based on Munoz et al. [MBRC15].

The publication by Kinkeldey et al. [KFBI21] used 8–axes star glyphs to visualize data of clusters and individual entities. These star glyphs are a type of star chart, that allows data on multiple axes and without the use of too much space. A disadvantage of this visualization is the amount of visualizations when selecting hundreds of participants and displaying a star chart for each participant. Because of this, the publication capped the number of star glyphs on one page at 400 [KFBI21].

2.2.3 Scarf Plot

Scarf plots are used in the eye tracking field to find Areas of Interest (AOIs) on gaze data or other types of data. Because the timelines for the displayed data are synchronized, the AOIs each different participant looked at can be seen in each specific moment of the timeline [KHW14]. With this, the AOIs for different viewers can be easily compared and the differences analyzed.

The scarf plot is used in VASD to visualize the differences in accelerometer data for multiple participants or for a single participant on selection. With it, the differences in daily activity should be definable and help in gaining insights about the variations of physical activity across groups of participants or the daily routine of a participant.

The book by Holmqvist et al. [HNA+11] presents various methods and measures to use in eye tracking and the paper from Kurzhals et al. [KHW14] presents a system called ISeeCube that shows multiple coordinated views to support visual analysis tasks on videos. AOIs can be more easily annotated and investigated with it [KHW14]. One view used in the system was the timeline visualization, which visualized temporal information of annotated AOIs and used scarf plots for individual viewers [KHW14].

In the paper by Blascheck et al. [BSBE17] an application is presented, which shows AOIs for eye movements. To compare the eye movement data of two or multiple participants and find AOIs a scarf plot was used, an example of the scarf plot can be seen in Figure 2.5.

To identify AOIs in the MoMo accelerometer data, the use of a scarf plot seems reasonable. With the scarf plot, patterns in daily activity for one participant or across a group of participants are identifiable.
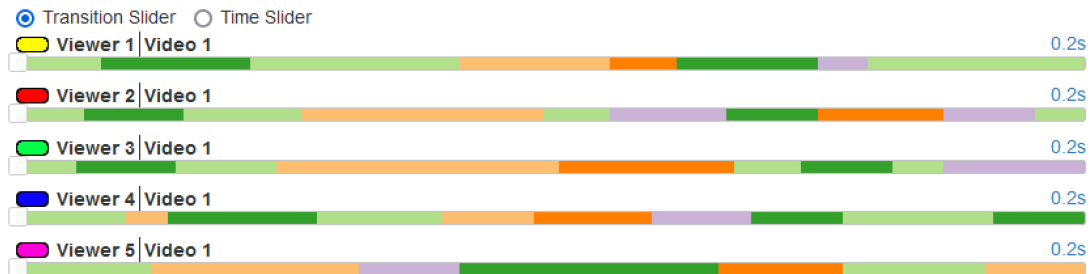
Figure 2.5: Scarf plot for multiple viewers, similar to figure in paper from Kurzhals et al. [KHW14], generated with the website Radial Transition Graph Comparison [SEBB].

In the publication by Yang and Wacharamanotham [YW18] the scarf plot was further developed to help with finding AOIs patterns which are distributed across multiple participants, as this is a problem for the familiar scarf plot.

2.2.4 Scatter Plot

The paper by Keim et al. [KHD+10] presents new ideas for improving scatter plots, including a discussion of the familiar scatter plot. Scatter plots „are one of the most common ways to visualize multidimensional data" [KHD+10]. With scatter plots relations between two attributes can be compared, clusters can be found, and extreme values recognized, see a scatter plot example in Figure 2.6.

To display multiple time waves of data for the various categories of VASD (Fitness, BMI and Questionnaire) and their relations between two selected data variables, a scatter plot is a good way to visualize this. One data variable is shown on the Y–axis and another one on the X–axis depending on what should be visualized. With the scatter plot huge amounts of participant data can be displayed in one chart, but an overlap of data points can hinder finding insights [KHD+10].

Figure 2.6: Scatter plot example of a relation between BMI and the Amount of push–ups done in one minute.
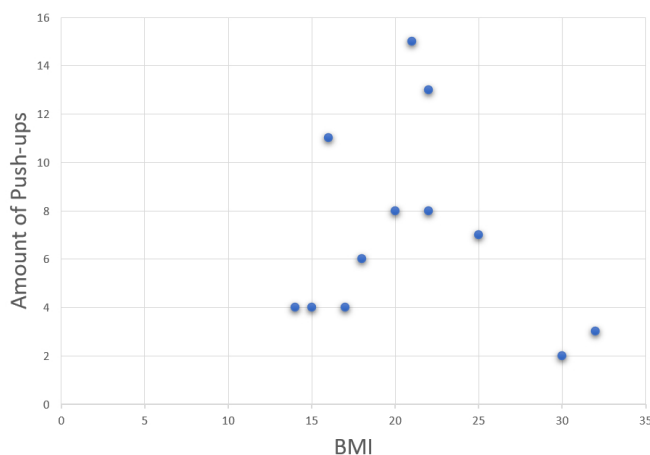
In the paper by Heer and Shneiderman [HS12] various visualization systems are presented, and one system called Spotfire uses a scatter plot to visualize the relation of the total money earned to the number of purchases for the genders female and male. This offers a fast way to see the fluctuation and distribution of the data set.

A Scatter plot was also tested in the publication of Section 2.2.1 [SED18], performing well in regard to accuracy and time to find correlations in the data. Additionally, it was useful in finding anomalies in the data set. The performance of the scatter plot in different tasks was always in the top three.

2.2.5 Glyphs on Maps

With the use of a geographic map and different types of Glyphs (markers, symbols) an overview of points of interest within the data is visualized. To enable a selection of different places of origin in VASD, a geographic map with interactable elements is a first overview for the data available. With the overview and interaction on the map, the selection of specific areas to start an analysis is a fast way to get new insights.

In the paper by Krüger et al. [KTW+13] a system called TrajectoryLenses is presented. In this system a geographic map is used to display different road paths, that electric scooters used through Stuttgart for a long time period (one year), see Figure 2.7. With the use of one or multiple lenses a more detailed analysis of all the areas inside a lens are possible.

The paper by Franke et al. [FMKK21] presents a system that contains a geographic map, which visualizes different types of events, for example earthquakes. The map contains a legend which explains the different sizes and colors of the circles on the map. These circles visualize the appearance of the events and on a selection of one or multiple of these circles, more details about their data is shown beneath the map.
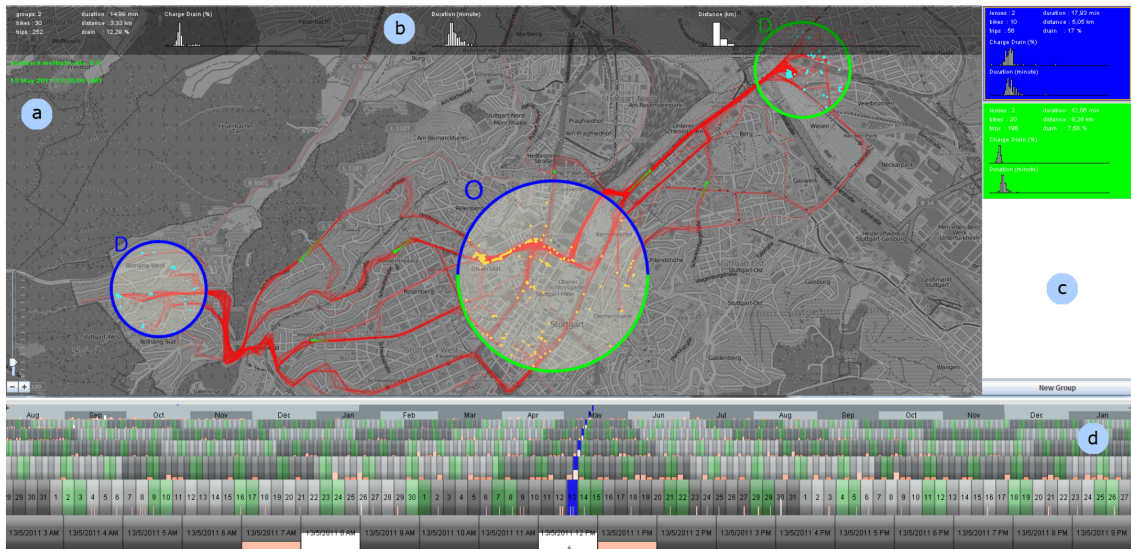
Figure 2.7: Overview on TrajectoryLenses [KTW+13] that visualizes hotspots of electric scooter routes. Figure reprinted with kind permission from Wiley Online Library [Wil].

In the publication by John et al. [JRK+17] a visualization system is presented, that uses a map view, to get an overview of the places where the data was published. One drawback of this type of visualization is the amount of space that can be empty when there is no data available for parts of the map.

2.3 Visualization Systems for Different Types of Data

The following subsections show different approaches to creating a visualization for various types of data. Each of the systems has multiple views to present the data set from different angles and tries to highlight these aspects of the data for the user, to gain insights or understand the underlying data better. Useful features of these systems are also used in VASD, for example multiple data filters or the ability to display different types of category data.

2.3.1 PhysioLab

The system PhysioLab [MBRC15] shows interactive information about a wearable electrocardiograph sensor. With this sensor the fitness of a user can be diagnosed, a part of the measurements is done while exercising. Parameters are extracted from the Electrocardiogram (ECG) data, for example maximal oxygen uptake or heart rate (Beats per minute (Bpm)) [MBRC15]. These parameters are then visualized with the use of line charts and scatter plots to show multivariate physiological data of the participant [MBRC15], see Figure 2.8.

Various filters are available to change and set different parameters of the program, for example the minimum and maximum values of the data window.
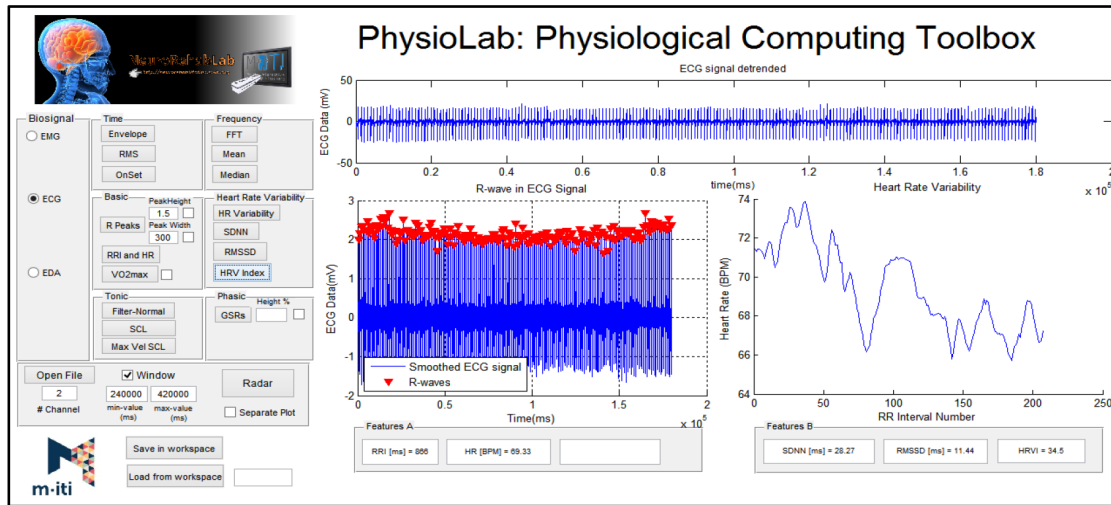
Figure 2.8: Overview of the PhysioLab system [MBRC15]. Figure reprinted with kind permission from Institute of Electrical and Electronics Engineers (IEEE) [EE].

With this system, an analysis of the participant's fitness status can be done, by comparing the ECG data to other referential data [MBRC15]. Past ECG measurements of a participant could be used to compare it to the current data set and see if the participant made progress in increasing their fitness. If there is no past data available, the current data set could be used to compare it with future measurements.

2.3.2 Fitster

Fitster [AGPR06] is a social fitness information visualizer which was created to help people with their struggle to do workouts. The system supports „activity tracking, goal setting and motivation through virtual competitions and teamwork" [AGPR06].
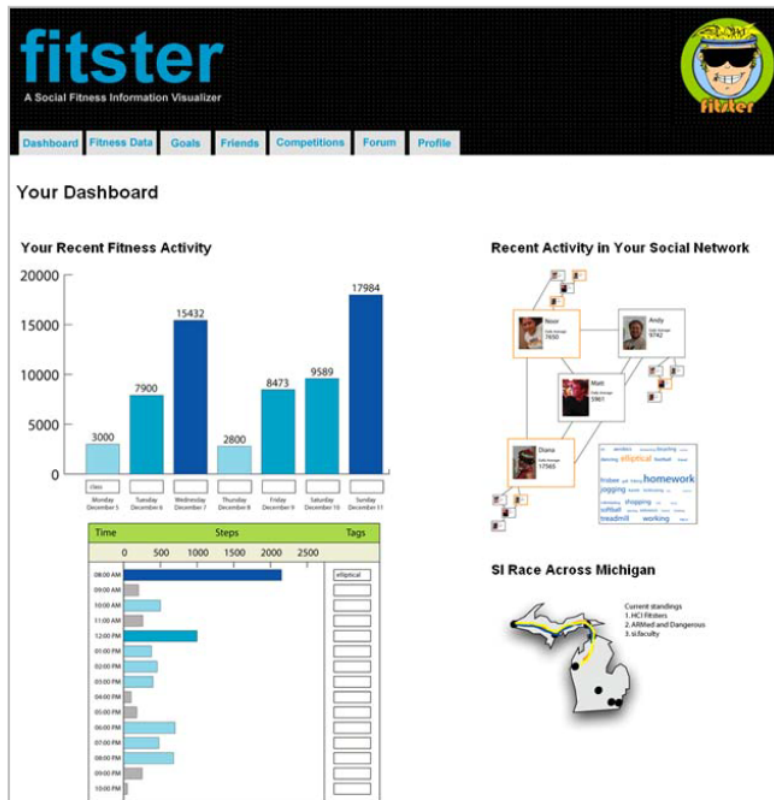
Figure 2.9: Overview of the Fitster system [AGPR06]. Figure reprinted with kind permission from Association for Computing Machinery (ACM) [Com].

There are different views available in Fitster, for example Dashboard, Fitness Data, Goals, and a Forum. In Figure 2.9 the dashboard is shown, where step count data is displayed in a bar chart, recent activities of friends and a geographic map with a race are also displayed [AGPR06]. Another detailed chart available is the line chart, which compares the step counts over two weeks for four participants. With this, the participants can compete and see others progress with their fitness goals.

2.3.3 Regionalatlas Statistics Portal

Regionalatlas [Bunb] is a system that uses a geographic map to visualize selected category data. There are various sectors (Building and living, Education, Tourism and many more) available and for these sectors, there are also different types of categories available, for example „Share of area for forest in total area" or „Proportion of sports, leisure and recreation space" which is shown in Figure 2.10. A legend is available for the map, which explains the visualized data and information on the topics.
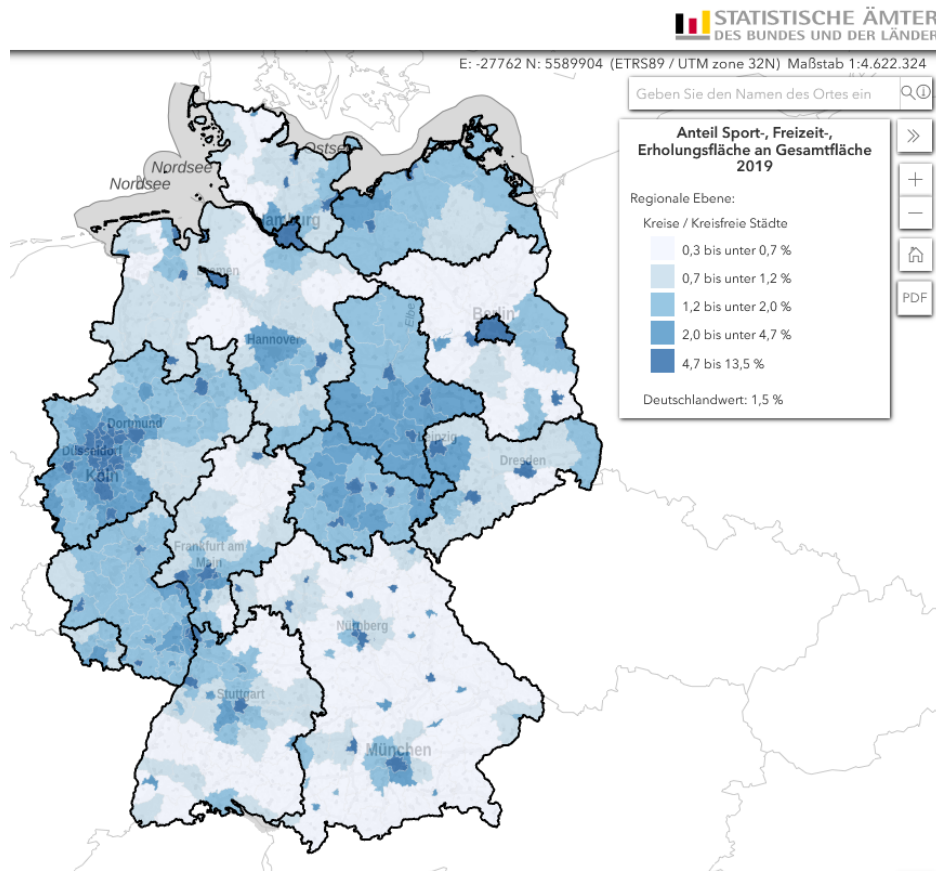
Figure 2.10: Excerpt from Regionalatlas [Bunb] about proportion of sports, leisure and recreation space in 2019, with permission „Datenlizenz Deutschland – Namensnennung – Version 2.0" [Gov] for the data set [Buna].

The system has filters to set the year of the displayed data or the area level that is shown, for example, federal states or community level.

2.4 Distinction to This Thesis

In PhysioLab (see Section 2.3.1) other fitness data (physiological data) is shown compared to the VASD system with more general fitness data for each participant. The focus of VASD is on analyzing a group of participants and if needed single participants, in PhysioLab the data analyzed is only for one participant.

Compared to Fitster (see Section 2.3.2), VASD offers an overview over many participants from the different time waves of the MoMo study, while Fitster is about individual users and their friends. In addition, VASD has multiple linked views, which change their displayed information on selection and the use of filters.

VASD offers a similar overview on the data set for different categories as Regionalatlas but improves on this further. Regionalatlas only shows the exact values behind the visualization on selection, while VASD offers other more detailed visualizations to analyze the data with them.

VASD has a data set that does not change quickly, for this a new wave of the MoMo study has to take place. In contrast, PhysioLab and Fitster can quickly add new data, for example Fitster updates data daily when users add their fitness workouts.

PhysioLab and Fitster are targeted at the analysis of individual users, while Regionalatlas and VASD are targeted to give an overview of the data sets. In addition, VASD supports the visualization of the data for single participants.

VASD is realized as a website that can be hosted in the cloud, the same as Regionalatlas, to allow for the general public to have access to the MoMo data. Because of this, the distribution of VASD to new users is as simple as sending a link to the website it runs on instead of installing a program.

3 Concept and System Implementation

In this chapter the concept of the VASD system is presented and all the steps needed to realize the system. For example, the processing of the MoMo data set, multiple algorithms or formulas used in the system are explained and the system design to realize all the visualizations.

3.1 Concept of the VASD System

A goal of the VASD system was to offer different views for the underlying MoMo data and to filter and extract more detailed information with a few clicks. For various parts of the analysis, there are different categories to choose from and for each category, there are multiple charts that show different parts of information. With the various data sets available, one requirement for the VASD system was to combine them and allow an analyst to analyze these data sources together, to allow linked analyzing. Some questions that should be answered with the system are, if there are connections seen within the data visualized or if there is anything other to be taken from the visualizations, for example new insights or findings. More detailed questions regarding the visualized data are:

- Does activity and motor performance differ by age or gender?

- Are there regions in Germany where children have a higher motor performance?

- Do motor performance and activity behavior differ across the different measurement waves?

- Can different progressions throughout the day be seen?

- Are motoric fit children more often active in a sports club?

- Is a child with playgrounds in the neighborhood healthier (lower BMI)?

- Does the way to school influence the fitness of children?

3.2 Used Tools and Libraries

Python [VF] version 3.7.1 is used as the back–end of the system for all necessary data provisioning and processing tasks, for example processing of all provided MoMo data sets to make them easier to use in the system when needed and in an already processed state. The Flask package [Pal] from Python is used on the webserver for the serving of the Application Programming Interface (API), which responds to different calls to supply the system with the requested data. For the front–end running on the webserver and handling the presentation of the website, React [WM] and Node.js [DF] version 14.15.1 are used in combination with JavaScript [EN] and npm [Sn] version 6.14.8.

Visualizations were realized with the use of Leaflet [Aga] for the geographic map visualization including its functions, in combination with D3.js [BDHO] version 5 for the group visualization. All the other charts seen in VASD are realized with D3.js.

3.3 Data Source and Processing of the Data

Multiple data sources were provided by the KIT, Statistical Package for the Social Sciences Statistics Data File Format Family (SAV) files contained all the participants anonymized data of the MoMo project, Microsoft Access Table Shortcut (MAT) files contained the tracked accelerometer data of participants and an Excel file was provided for the variables, labels, and values to use the SAV files. In addition, an Excel file for the mapping of the accelerometer data to the participants data was provided and with this, linking between these data sources is possible. The general approach was to integrate the data from the files as generically as possible, reading in the variables, labels, and values and mapping them to the available data. This was possible for almost every category and subcategory. Some additional mapping needed to be done for the „Schulweg" parameters, for more information see Section 3.3.4. The processed data of participants was saved in SQLite [Hip] databases and their linked accelerometer data in Parquet [Apa] format files, which were separated by day to allow for faster loading times.

3.3.1 Privacy

On first integrating the positions of participants into the geographic map, the location information was taken from the provided postal codes of the questionnaire. With these postal codes the specific region where a participant lives could be displayed, but this emerged as an issue for participants which lived in cities or villages without any other participants around them because then they potentially could be identified. To ensure the privacy of these participants, this detailed home information was aggregated into the state in which participants live. Because of this the participants are all shown in one of the 16 state groups, which makes it near to impossible to identify participants based on their location.

3.3.2 Accelerometer Data Processing

The provided accelerometer data is saved in 15 second epochs, in total there is data of 1,880 participants available. The 15 second epochs of accelerometer data are provided as counts indicating how many accelerations the device measured. These count values are fed into the Choi algorithm to decide on wear and non–wear times. After the choi algorithm these processed files are categorized with Evenson or Romanzini cut–points to divide them into sedentary, light, moderate and vigorous activity [BMA+20], for exact values to categorize the data see Table 3.1. The data collected is from time wave T3, where participants wore the accelerometer device to record eight days straight which physical activities they do during the day.

Data Cleaning with the Choi Algorithm

The choi algorithm [CLMB11] is divided into three parts:

- The algorithm has two time windows for which the data is checked, the first window is 90 minutes long, in which all the 15 second epochs are consecutive zero counts.

- The second window allows for short intervals of up to 2 minutes with non–zero occurrences, if before and after this non–zero interval there are 30 minutes of consecutive zero counts.

- If time windows are found which fit the described windows, these intervals of the data are marked as „Non–Wear" time. Everything else in the data is considered as „Wear" time.

The wear time must be longer than 8 hours for each day, otherwise the day is not used, because it would not be representative of daily use.

Categorize Data with Evenson or Romanzini Cut–points

The MoMo study used the cut–points from Evenson et al. [ECG+08] and Romanzini et al. [RPO+14] to classify the wear counts of the accelerometer data. These cut–points and the age ranges are also used in this thesis. For the age range 6 to 10 the Evenson cut–points are used and for the age range 11 to 17 the Romanzini cut–points. The values of these cut–points were originally defined as counts per minute (counts * min^{-1}), but because the accelerometer data is in 15 second epochs, the values of the cut–points were divided by 4, which results in counts per 15 seconds (counts * $15sec^{-1}$), as seen in Table 3.1.

Table 3.1: Cut–points defined by Evenson and Romanzini (counts * $15sec^{-1}$).

Cut–points name	Participant age	Sedentary	Light	Moderate	Vigorous
Evenson	6–10	0–25	26–573	574–1,002	>=1,003
Romanzini	11–17	0–46	47–606	607–817	>=818

After the processing steps with the choi algorithm and the Evenson and Romanzini cut–points with the age range 6 to 17 and eight hours of wearing per day, the processed accelerometer data of 1,718 participants remained.

A processed accelerometer file would look like the excerpt of Figure 3.1. The first column „Datetime" is in ordinal format and contains the date and time when a 15 second epoch occurred. The second row „Count" contains the count that was measured for this epoch, the third column „Cutpoints" contains the divisions which were decided upon the cut–points and the fourth column „index" contains the index to help with processing in Python.

Datetime	Count	Cutpoints	__index_level_0__
1481614170000000	0	Non Wear	24838
1481614185000000	0	Non Wear	24839
1481614200000000	0	Non Wear	24840
1481614215000000	0	Non Wear	24841
1481614230000000	0	Non Wear	24842
1481614245000000	0	Non Wear	24843
1481614260000000	0	Non Wear	24844
1481614275000000	0	Non Wear	24845
1481614290000000	474	Light	24846
1481614305000000	14	Sedentary	24847
1481614320000000	25	Sedentary	24848
1481614335000000	0	Sedentary	24849
1481614350000000	30	Sedentary	24850
1481614365000000	3	Sedentary	24851
1481614380000000	0	Sedentary	24852
1481614395000000	50	Light	24853
1481614410000000	97	Light	24854
1481614425000000	0	Sedentary	24855
1481614440000000	119	Light	24856
1481614455000000	107	Light	24857
1481614470000000	112	Light	24858
1481614485000000	229	Light	24859
1481614500000000	104	Light	24860
1481614515000000	23	Sedentary	24861
1481614530000000	291	Light	24862
1481614545000000	151	Light	24863
1481614560000000	252	Light	24864
1481614575000000	290	Light	24865
1481614590000000	435	Light	24866
1481614605000000	63	Light	24867
1481614620000000	0	Sedentary	24868
1481614635000000	558	Light	24869
1481614650000000	398	Light	24870
1481614665000000	0	Sedentary	24871
1481614680000000	385	Light	24872
1481614695000000	332	Light	24873
1481614710000000	240	Light	24874
1481614725000000	73	Light	24875
1481614740000000	197	Light	24876
1481614755000000	362	Light	24877
1481614770000000	276	Light	24878
1481614785000000	236	Light	24879
1481614800000000	217	Light	24880
1481614815000000	158	Light	24881
1481614830000000	225	Light	24882

Figure 3.1: Datetime, count, cut–points division and index columns of a processed accelerometer file.

3.3.3 Processing Fitness Specific Data

For the different time waves available in the data set

$$T_i \in \{1, 2, 3, 4, 5, 6\}$$

the following formulas are used to calculate scores, that are displayed when the category Fitness is selected. Values behind the parameters are represented in percentiles, $sitPerzA\_T_i$ (amount of sit–ups in 40 seconds) was not covered in time wave T1 and $PWC170PerzA\_T_i$ (static bicycle ergometer test at a heart rate of 170 Bpm) only was covered if the participant was older then 9 years old [NUO+20]. Because of this some formulas first resulted in „Not available," but on viewing the details of single participants scores were still available which caused confusion. Therefore, a modification of the formulas was made.

If one of the parameters is missing (was not collected for the time wave), the result gets calculated with the parameters which are available and if there is a division in the formula, it is divided by the number of available parameters. If all the parameters are missing, the score gets categorized into „Not available," which means the participant participated in the time wave but there is no value available. „Not participated" is used if the participant did not participate in the time wave, for which the formula is calculated. If one of the parameters is missing for the calculation of the total score fitness formula, this is marked for the participant and displayed as additional information in the form of a texture and a tool–tip in the visualization.

Because the parameters used in the formulas are abbreviated, here are the long descriptions of the parameters, everything is classified in percentiles according to Niessner et al. 2020 [NUO+20][WWA+15]:

- **weitmaxPerzA:** Standing long jump, jumping as far as possible from a standing position, maximum from two attempts.

- **liegPerzA:** Amount of correct push–ups performed in 40 seconds.

- **sitPerzA:** Amount of correct sit–ups performed in 40 seconds.

- **seitlxPerzA:** Amount of jumps while jumping sideways between two squares in 15 seconds, mean result from two attempts.

- **einbPerzA:** Single–leg stand with eyes open balancing forward on a beam without touching own leg or ground, number of mistakes counted in one minute.

- **balle6PerzA:** Amount of steps to balance backwards on three differently broad beams, all six attempts summed up.

- **PWC170PerzA:** The power produced at a heart rate of 170 Bpm was measured (Physical Working Capacity (PWC)) with a static bicycle ergometer test.

- **standmaxPerzA:** Stand and reach own toes with fingers, difference measured in centimeter, the best value taken from two attempts.

- **mlssPerzA:** Insert pins into a plate as fast as possible, two attempts.

- **reakxPerzA:** Respond to the green light as fast as possible, mean result of the best seven attempts.

All the different fitness formulas used are shown here:

Score Total Fitness

In this formula fitness relevant parameters are added up and divided through the number of parameters available, which returns a score that is declared as total fitness and describes the participant's fitness status across the different fitness disciplines.

$$\frac{weitmaxPerzA\_T_i + liegPerzA\_T_i + sitPerzA\_T_i + seitlxPerzA\_T_i + einbPerzA\_T_i+}{8}$$
$$\frac{}{balle6PerzA\_T_i + PWC170PerzA\_T_i + standmaxPerzA\_T_i}$$

Score Strength

This formula adds up all parameters of a participant which are related to strength, divided by the number of parameters and thus returns a strength score.

$$\frac{weitmaxPerzA\_T_i + liegPerzA\_T_i + sitPerzA\_T_i}{3}$$

Score Coordination

To evaluate coordination ability, all relevant parameters that are a measure of coordination tests are summed up and divided by the number of coordination parameters.

$$\frac{seitlxPerzA\_T_i + einbPerzA\_T_i + balle6PerzA\_T_i}{3}$$

Score Reaction Speed

This score shows how good participants react to optical attractions.

$$reakxPerzA\_T_i$$

Score Mobility

The participant's flexibility is measured and results in the mobility score.

$$standmaxPerzA\_T_i$$

Score Endurance

The power generated at a heart rate of 170 Bpm shows how good a participant's endurance is.

$$PWC170PerzA\_T_i$$

Score Fine Motor Skills

This score is about fine motor skills, testing participants on their eye–hand coordination.

$$mlssPerzA\_T_i$$

3.3.4 Defining „Schulweg"

When integrating the category Questionnaire and its various subcategories, there was a problem with the subcategory School route that occurred when processing the subcategory from the data set. Looking at the data set (see Figure 3.2), this subcategory had multiple parameters named Schulweg1 with a different amount of values for each time wave and for time wave T2 there were even Schulweg1 – Schulweg4. This made it necessary to summarize the values into a few divisions that match all available parameters and time waves.

Figure 3.2: Schulweg1 classifications found in the provided Excel file.

Table 3.2: Schulweg_T1 and Schulweg_T2 new mappings for the original Schulweg values.

Index	School route	Schulweg_T1	Schulweg_T2
1	On foot, by bicycle, long–board, running bike or scooter	1, 2, 11	1, 2
2	By bus or train	3	3
3	On foot, by bike, bus or train	9, 10	(1 or 2) and 3
4	By car, moped, motorcycle or scooter	4, 5	4, 5
5	Most means of transportation	6, 7, 8	(1 or 2 or 3) and (4 or 5)
6	Not employed, no studies or education	–	6

In Table 3.2 and Table 3.3 the mappings applied to the data can be seen. To receive one summarized Schulweg for time wave T2, the different Schulweg1_T2 – Schulweg4_T2 parameters had to be combined. For example, to receive Schulweg_T2 Index 3 in Table 3.2 with the mapping of (1 or 2) and 3, one or more of the parameters Schulweg1_T2, Schulweg2_T2, Schulweg3_T2 or Schulweg4_T2 have to be of value 1 or 2 and one of the remaining Schulweg1_T2 – Schulweg4_T2 parameters has to be value 3. Together this then results in the mapping of Index 3, which means „On foot, by bike, bus or train." To receive Schulweg_T1 Index 5, the old Schulweg1_T1 value has to be 6, 7 or 8, then the mapping results in „Most means of transportation", because at least two types of movement are used in the previous value of Schulweg1_T1.

In general, the subdivisions which were chosen can be distinguished as:

- personal movement

- public transport

- personal movement or public transport

- personal vehicles

35

Table 3.3: The new mappings of Schulweg_T3 and Schulweg_T4 from the original Schulweg values.

Index	School route classification	Schulweg_T3	Schulweg_T4
1	On foot, by bicycle, longboard, running bike or scooter	1, 2	1, 2
2	By bus or train	3	3
3	On foot, by bike, bus or train	–	–
4	By car, moped, motorcycle or scooter	4, 5	4, 5
5	Most means of transportation	–	–
6	Not employed, no studies or education	6	6

- combination of personal movement and vehicles and public transport

- not available records because of personal reasons

3.4 System Design

The main view with a geographic map on the left side (Section A) is always visible across the whole VASD system, for an overview see Figure 3.3 with category Fitness selected. This view gives an overview of the selected category data, using donut charts for groups or pie charts for single markers.
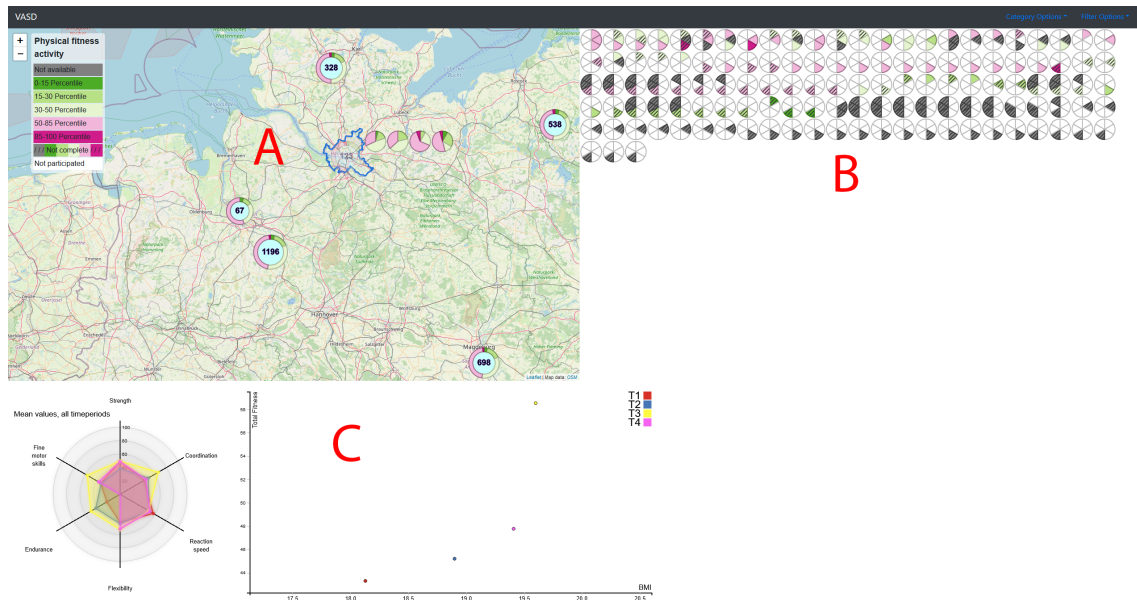

Figure 3.3: Total fitness overview with sections A, B, C. Geographic map with groups (A), participants pie charts (B), radar and scatter plot for more detailed information (C).

All the different donut charts of the four categories Accelerometer, Fitness, BMI and Questionnaire can be seen in Figure 3.4.
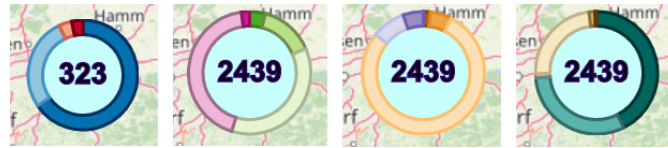

Figure 3.4: Accelerometer, Fitness, BMI and Questionnaire (School route) donut charts with the amount of participants displayed in the middle.

Many participants of a state are combined into a group, this was done to handle large amounts of participants (Figure 3.3 A). The groups are created from the data of each participant in the group, each time wave and filter selected. Each division with a color in the donut chart represents one of the divisions of the selected category and these divisions are filled up with data for each available time wave of the participants. The number in the middle of the donut chart represents the number of participants in this group.

An example to understand how the donut charts are created in the category Fitness: Data of two participants is available for the time waves T1, T2 and T3. One participant was categorized into the 85–100 percentile range and the other participant was categorized into the 50–85 percentile range for all three time waves. These data points are added to the donut chart and would show up with the tool–tips „50–85 Percentile (3)" and „85–100 Percentile (3)." The donut chart would have the divisions 50–85 and 85–100 percentile, both taking 50 percent of the space, see Figure 3.5. Time waves with no available data, because there was no participation or testing, are not added. Because the donut chart is generated across all participants, the chart shows the distribution over all the category divisions.
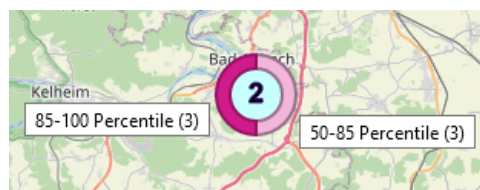

Figure 3.5: Example of a donut chart with two participants and the tool–tips „50–85 Percentile (3)" and „85–100 Percentile (3)."

In the current data set for VASD there are up to 13,445 data points for participants available, which then are aggregated by the different states. Clicking on a group shows more detailed charts on the right side (Figure 3.3 B) and beneath the map (Figure 3.3 C). On the right side (Figure 3.3 B) each participant who is inside the selected group is shown as a pie chart, with the category's data for this participant. The charts beneath the geographic map (Figure 3.3 C) also depend and change on the selected main category and at first show the mean average of the selected group.

3.4.1 Data Categories

There are four different categories to choose from, which display different data metrics. Category Accelerometer shows how active participants were during the day while wearing the accelerometer device. With category Fitness a total fitness score is displayed, which is classified into percentiles and consists of six different disciplines. The category BMI shows the distribution of the BMI for all participants. The category Questionnaire consists of various subcategories that reveal more about the participants environment. For category Accelerometer only time wave T3 is available, but for the categories Fitness, BMI and Questionnaire time waves T1–T6 are selectable individually and filter the data for the selected time waves.

If the category Accelerometer is selected, the chart shown beneath the geographic map (see Figure 3.3 C) is a scarf plot (Section 2.2.3) which details how active participants were during the whole day (00:00–24:00). With category Fitness two charts are shown beneath the map, a radar chart which shows in more detail the fitness disciplines and a scatter plot with a relation between Total fitness and BMI (see Figure 3.11). For the categories BMI and Questionnaire only the scatter plot is shown, for category BMI with the relation Days active per week doing sport & BMI (Figure 3.13) and for category Questionnaire the relation between Total fitness and the Questionnaire subcategory divisions (Figure 3.15).

Different color scales are used for each of the four categories, to not confuse the users of the VASD system when inspecting the data. With different colors, there is a strict distinction between the categories and therefore can not be misinterpreted while analyzing data. The colors used were selected with the online tool Color Brewer 2 [BHU]. The colors were chosen to be suitable for people with color blindness, and a divergent color scale was used to allow clear differentiation between the divisions.

Category Accelerometer

The accelerometer data, which is cleaned up as described in Section 3.3.2, is separated into four intensities of activity [BMA+20]. The four intensities with activity examples are:

- **Sedentary:** Sitting on a chair.

- **Light:** Cooking food at home.

- **Moderate:** Painting walls.

- **Vigorous:** Playing soccer outside.

With these divisions, the activities during the day can be distinguished as easy not demanding activities up to very demanding physical activities.

Category Fitness

The result of the total fitness score formula seen in Section 3.3.3 gets divided into five divisions, derived from Niessner et al. [NUO+20]:

- 0–15 Percentile

- 15–30 Percentile

- 30–50 Percentile

- 50–85 Percentile

- 85–100 Percentile

The other scores in the fitness category are about strength, coordination, reaction speed, mobility, endurance and fine motor skills of the participants. These formulas are described in more detail in Section 3.3.3.

Category BMI

With the category BMI, there are two subcategories to choose from: Body Mass Index according to Kromeyer-Hauschild (BMI KH) and Body Mass Index according to International Obesity TaskForce (BMI IOTF). Both subcategories have divisions with brackets and a P value in it. P means percentile and describes when a participant with a particular BMI is categorized into that division.

BMI KH has the following divisions:

- Severely underweight (<P3)

- Underweight (P3 – <P10)

- Normal weight

- Overweight, not obese (>P90 – P97)

- Obese (>P97)

BMI IOTF has the following divisions:

- Underweight (<P_BMI18.5)

- Normal weight

- Overweight, not obese (>P_BMI25 – P_BMI30)

- Obese (>P_BMI30)

Category Questionnaire

For the category Questionnaire the participants answered different questions about their way to school and their surroundings where they live, for example if there are playgrounds near their home. Therefore, there are different subcategories to choose from, the subcategories in addition to the questions and answer possibilities are [WKO+11]:

- **School route:** How do you usually get to school?

- **Sports facilities nearby:** In the area where I live, there are sports facilities freely accessible at all times. Answers: None, Few, Some, Many.

- **Sports clubs nearby:** There are sports clubs in my neighborhood. Answers: None, Few, Some, Many.

- **Commercial sports providers nearby:** In the area where I live, there are commercial sports providers. Answers: None, Few, Some, Many.

- **Playgrounds nearby:** There are playgrounds in my neighborhood. Answers: None, Few, Some, Many.

- **Sidewalks nearby:** To walk safely, there are sufficient sidewalks and walkways in my living environment. Answers: None, Few, Some, Many.

- **Bike lanes nearby:** There are bike lanes in the area where I live. Answers: None, Few, Some, Many.

- **Cars driving nearby:** In my living environment cars drive. Answers: None, Few, Some, Many.

- **Children play sports nearby:** In my living environment other children and young people play sports. Answers: None, Few, Some, Many.

- **Safety of public recreational facilities nearby:** How safe are the public recreational facilities in your neighborhood? Answers: Very uncertain, Rather uncertain, Rather safe, Very safe.

- **Walking and cycling nearby:** For walking and cycling my living environment is... Answers: Very unpleasant, Rather unpleasant, Rather pleasant, Very pleasant.

- **Stores nearby:** In the area where I live, stores are within walking distance. Answers: Very poorly accessible, Rather poorly accessible, Rather well accessible, Very well accessible.

- **Bus and streetcar stops nearby:** From my home the bus and streetcar stops are within walking distance. Answers: Very poorly accessible, Rather poorly accessible, Rather well accessible, Very well accessible.

For the subcategory „School route", a mapping had to be made because time wave T1 and T2 had more than the usual six parts of the other time waves. This new mapping is described in Section 3.3.4.

3.4.2 Filter Data

Seven different filters are available to filter the data, these are:

- Age

- Sex

- City size

- Social status

- BMI

- Migration background

- Sports club member

- Participant IDs

For the filter Participant IDs the input is comma separated, for example „88800060, 88800061"
would be a valid input. Filter Age only accepts number inputs without comma separation, but has
multiple input variants with „<=, >=, – (range) and =" as a selection. All filters are combineable
with each other and can be used to filter to small sub samples of the data set. Additionally, the
selection of different time waves helps to filter for specific use cases. These category, time and filter
selections are available on the top right side of the system under filter options and category options.
Overview of the selection options in Figure 3.6.

Figure 3.6: Category and subcategory selection, selection of time waves and all the available filters.

Under filter options, there is a date range option for the category Accelerometer which contains
a legend to see how many participants have accelerometer data for each day, see Figure 3.7. The
distribution of the legend was selected to divide the amount of available accelerometer data into four
parts. It ranges from under 9 participants of data for a day, to over 24 participants with accelerometer
data on the specific day.

A single day can be selected to receive data for this day only, or a date range to receive multiple days worth of data. For the categories Fitness, BMI and Questionnaire a month range option is available, which offers the same functionality with selecting single or multiple months, see Figure 3.7. On selection of one month or multiple months, only participants that participated in the selected months are shown, the year is not taken into account. Selected days or months are highlighted in blue.
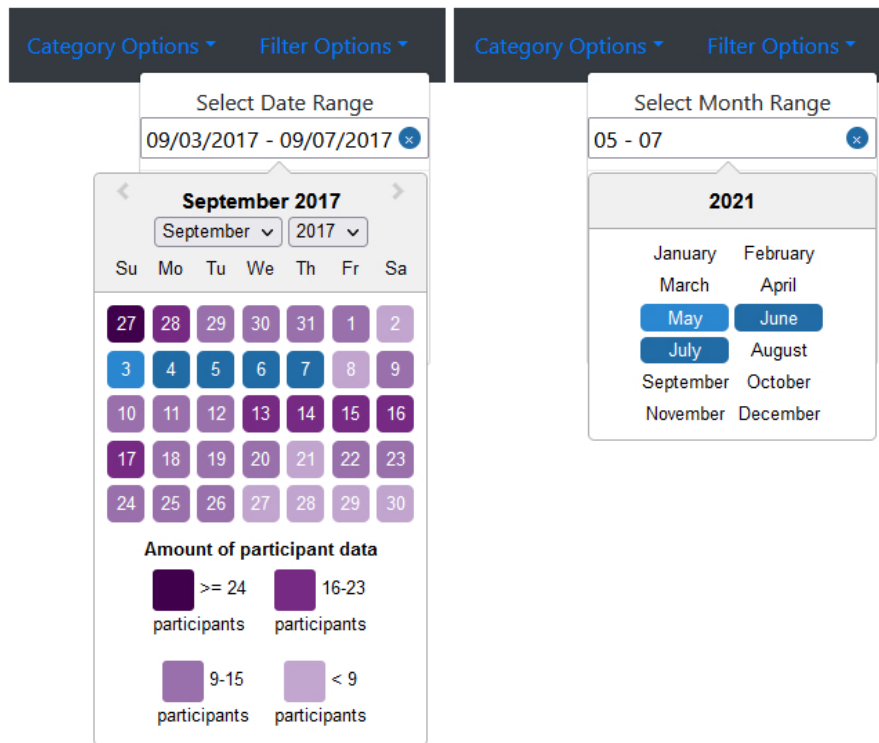

Figure 3.7: Date range selection for category Accelerometer and month range selection for categories Fitness, BMI and Questionnaire available.

3.4.3 The Different Category Views

In this subsection, the four VASD category views are presented. Each category has different charts available or the same charts with different data variables selected. Only the geographic map is always visible to show where participants are located. Combined with the donut charts shown on the map, it offers a good overview, to see how the selected category data is distributed. Pie charts for each participant are displayed after the selection of a group or group time wave, to show the underlying data for each participant.

Accelerometer View

The click on a group on the geographic map, selects all the participants to show their scarf plot
with the accelerometer data, which fits the participants, the date range and shows each day found,
ordered after the dates of the data. Due to the long loading times, only the first ten participants of a
group are shown in the scarf plot. Clicking on one or multiple of the pie charts which appear on the
right side after selecting one group, selects these participants to show their data on the scarf plot
instead. The scarf plot allows for visual tracking of the physical activity and leads to analysis of
the accelerometer data, see Figure 3.8. For an overview about the activities in a day the different
intensities of motion (Sedentary, Light, Moderate, Vigorous) are distinguished with different colors.
There is also the option to select a smaller time span, instead of 0–24 o'clock, there are choices for
6–24, 7–13, 13–18 and 18–24.
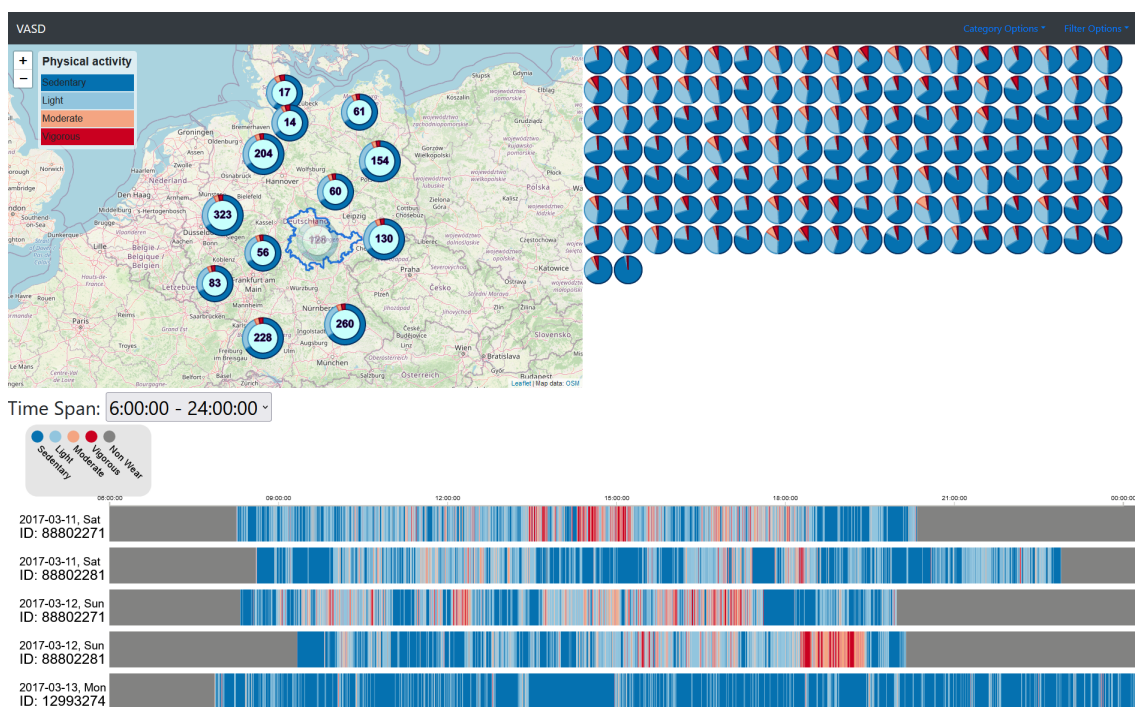
Figure 3.8: Overview of category Accelerometer with donut and pie charts and the scarf plot with
accelerometer data.

The goal of the Accelerometer view is to show the different activity patterns, participants had while
wearing the accelerometer device. With it, the user should have the possibility to see if participants
were active or did not do any strenuous activities during the day. Hovering over parts of the scarf
plot shows more information about the time this specific part of activity happened and how long it
lasted in seconds, see Section 3.4.3.

Fitness View

With the Fitness view, the user of VASD gets an overview of the fitness status of the participants. Clicking on a group allows for a more in depth analysis of the fitness activities, see Figure 3.9. The pie charts on the right side are divided into the number of time waves selected, to capture different states of fitness and health of the participants. If all six time waves are selected, this means that the pie chart is divided into six slices. Each of the slices represents one of the time waves in numerical order clockwise. 85–100 percentile are the participants with the best scores in these fitness tests. With this pie chart, the fitness status can be checked to see if there was a change over the different time waves, in case the participant is a longitudinal subject. If a participant did not participate in a time wave the slice stays white with tool–tip „Not available." If the participant took part in the time wave, but not in the fitness activity tests, the slice is gray, with tool–tip „Not participated."
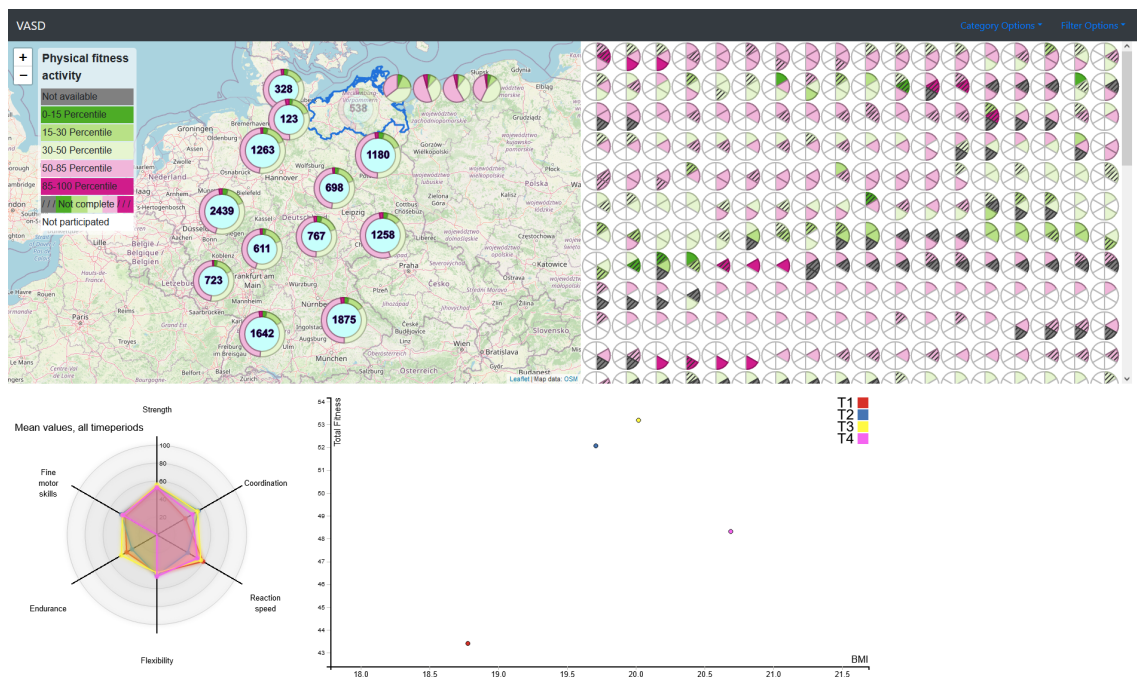

Figure 3.9: Category Fitness overview with the donut and pie charts and the more in detail radar and scatter plots.

If one or multiple of the fitness tests are missing because the participant did not take part in it, but the total fitness score was calculated with the leftover parameters, the slice has the normal categorization, but in addition, is marked with a texture. This should signal the user, that a total fitness score was calculated, but some parameters for the calculation were missing. If the user hovers over this specific crossed out slice, the tool–tip tells the user which parameters were missing in the calculation.

On selection of a single participant in the pie charts, the radar chart shows the fitness disciplines over the different time waves for this participant alone. With this, the performance of the participant can be analyzed, for example if they got better or worse in some tests. While hovering over one

of the areas colored, the selected time wave is displayed to help with orientation, see Figure 3.16. With this more detailed view, the different time waves can be compared and checked, if there is some significant difference between them.

The scatter plot shows the relations of Total fitness to BMI across the different time waves. If more than one participant is selected in the pie charts, a radar chart is shown for each of the participants with their performance in the fitness disciplines and the scatter plot shows the different data points for all selected participants and time waves together, see Figure 3.10.
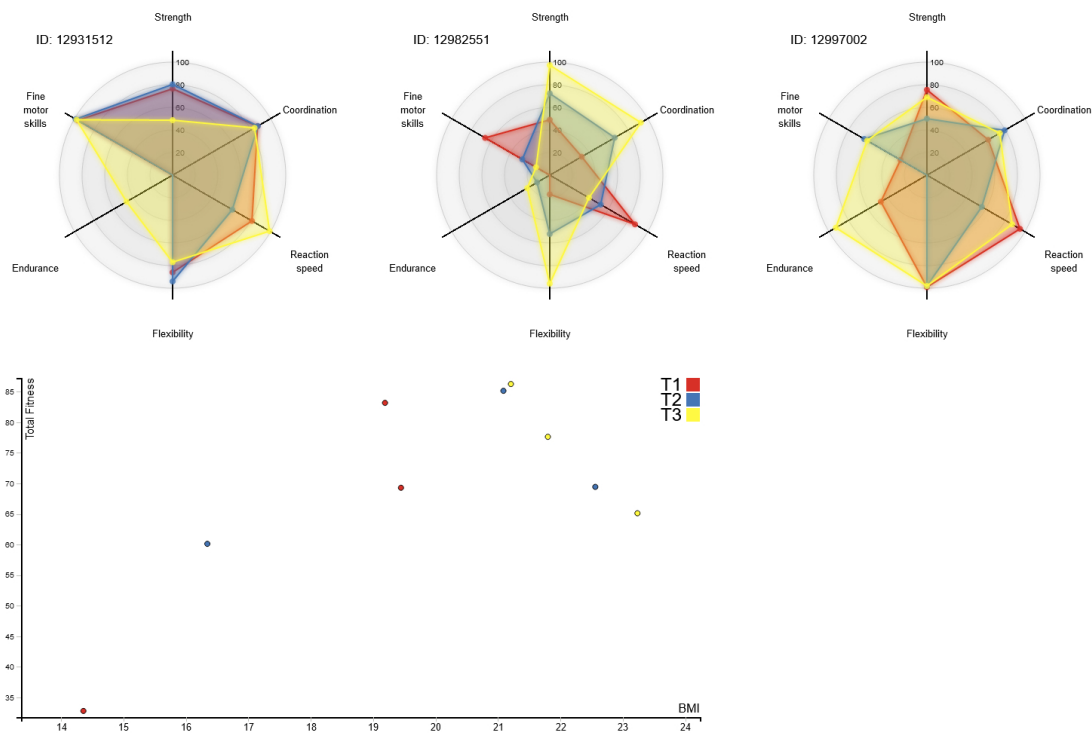
Figure 3.10: Category Fitness selected participants in detail analysis with the use of radar and scatter plots.

If one of the time waves beside the clicked group is selected, again the radar chart shows the average of the whole group, but this time only for the selected time wave, see Figure 3.11. The radar chart has the same values as the click on the group, but it helps with analyzing time waves separately. The scatter plot shows all the participants which took the study in the time wave with their values.
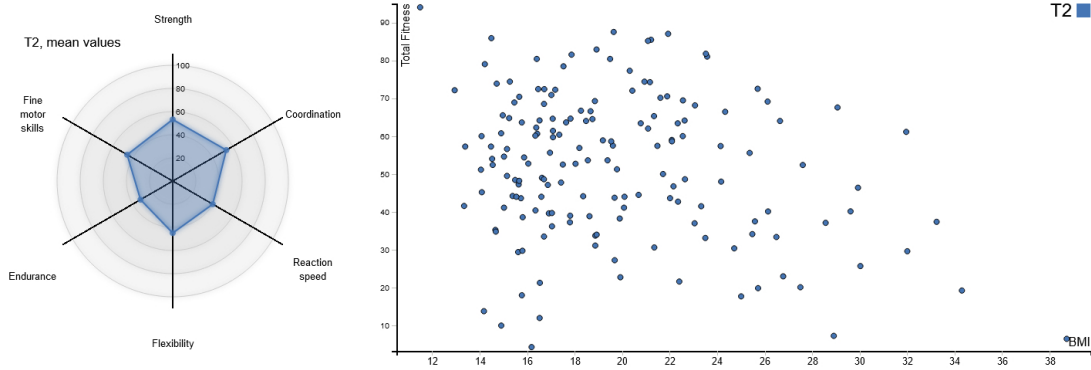
Figure 3.11: Fitness analysis for state Mecklenburg–Vorpommern and selected time wave T2 in detail.

BMI View

With category BMI, the user gets an impression of the overall BMI distribution, overview in Figure 3.12. There are two subcategories to choose from, BMI KH which is preselected and BMI IOTF. On the pie charts, the user sees for each time wave the selected BMI categorizations a participant had over the different time waves. The scatter plot shows a relation between the days the participant was actively doing sport in a week and the BMI the participant had in a time wave.
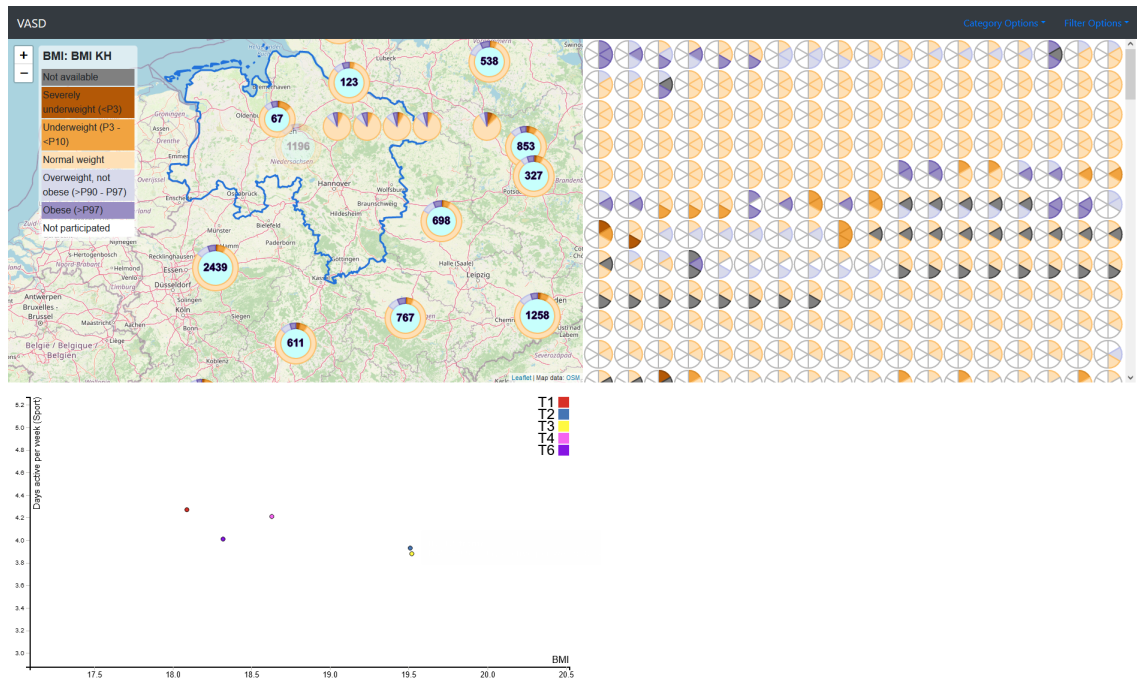
Figure 3.12: Category BMI overview with selected donut chart, pie charts and scarf plot for state Niedersachsen.

Selection of a specific time wave on the group again shows the scatter plot with all the participants who participated in this time wave, see Figure 3.13. Selecting one or multiple pie charts shows the relations of these participants data in the scatter plot.
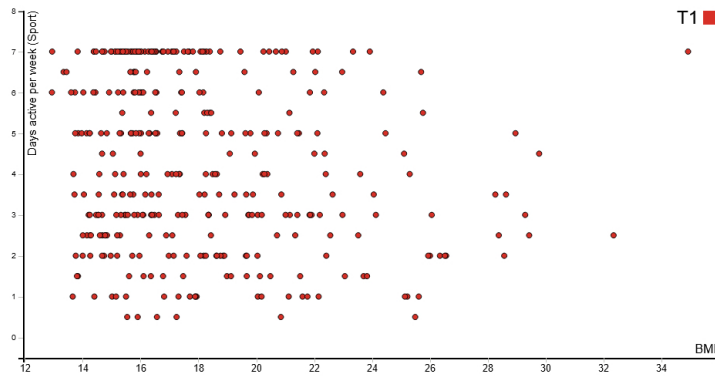
Figure 3.13: BMI scatter plot for all selected participants in Niedersachsen that participated in time wave T1.

Questionnaire View

In the category Questionnaire there are different subcategories to choose from, these are listed in Section 3.4.1. The subcategory preselected for this section is School route, which shows how children and teenagers travel to school, see Figure 3.14. For each subcategory the legend of the map updates accordingly.
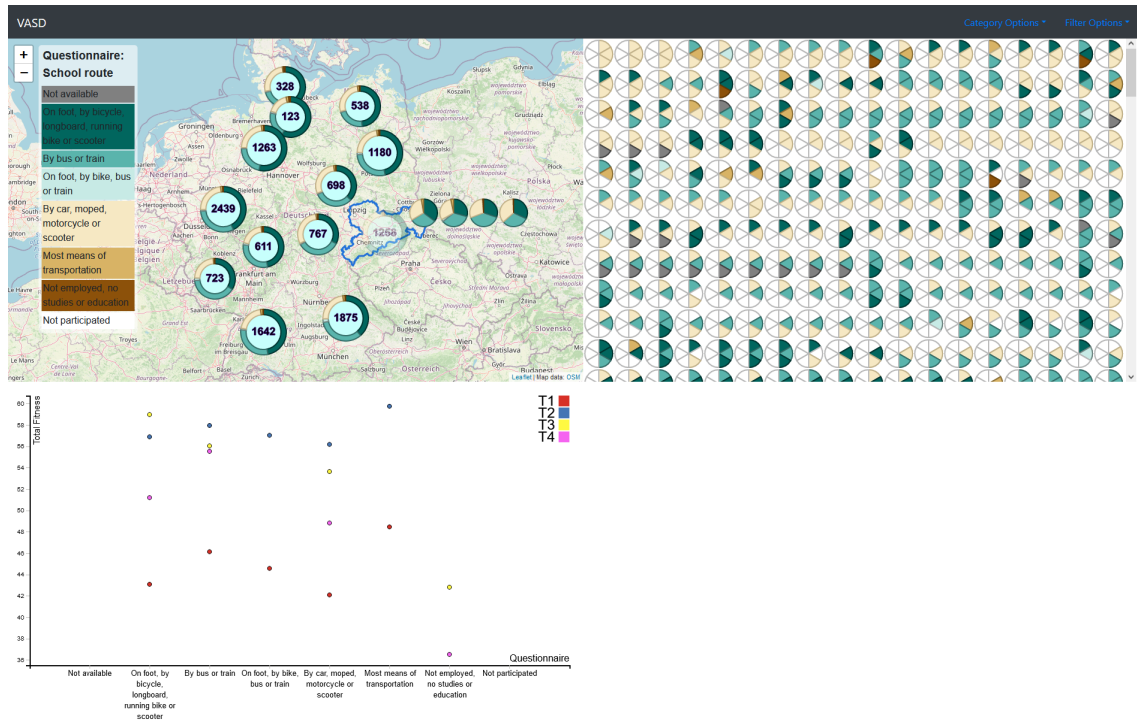
Figure 3.14: Overview on category Questionnaire (School route) with state Sachsen selected.

Again, the pie charts show the different values of how the kids got to school for each of the selected time waves. On selection of a group, group time wave or single participants, the scatter plot shows the relation between the total fitness of the participants and the selected way to school (see Figure 3.15). For each of the divisions of the Questionnaire subcategory, the answers are shown for the available and selected time waves. With this, a direct comparison of the relation between the total fitness and the chosen subcategory is seen for the time waves. For a group, this is again averaged and for group time waves all participant relations are shown (Figure 3.15).

Figure 3.15: In detail scatter plot analysis of all selected participants for time wave T3.

Details in the VASD System

While hovering over a participant's pie chart there is additional information about the participant seen. Their Identifier (ID) used in the MoMo study, their age in different time waves they took part in, their gender, the time wave 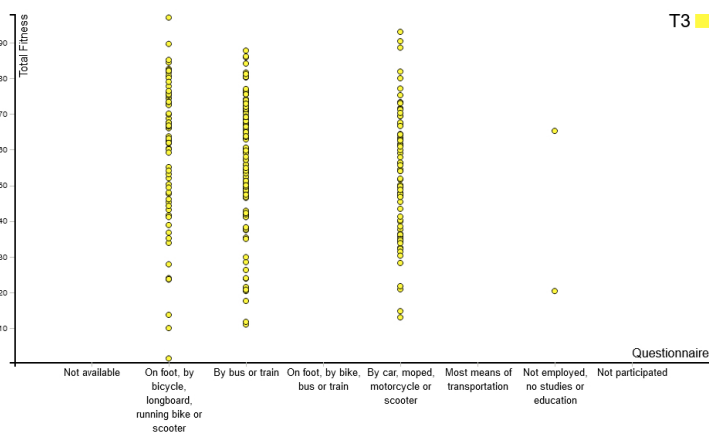of the studies and the state where they live. Also, for the selected category or subcategory, the values that form the pie chart become visible when you move the mouse cursor over the slices of the pie chart, see Figure 3.16.
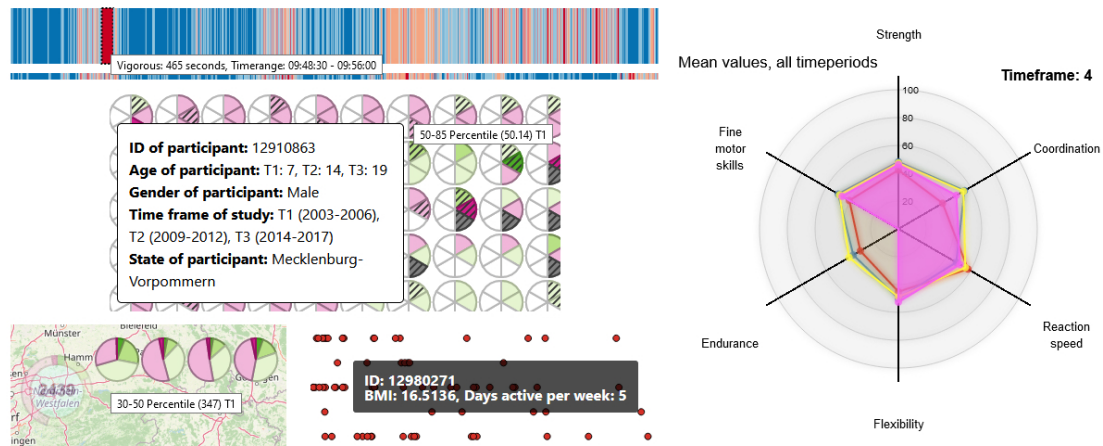

Figure 3.16: Tool–tips for participants, shown on scarf plot, pie chart, scatter plot and radar chart.

In the radar chart, the values for the whole group time waves or specific participants are shown for each category tested. To help with orientation, while hovering over one of the colored areas, the corresponding time wave is displayed, see example in Figure 3.16.

3.5 System Workflow

The workflow of VASD shown in Figure 3.17 is inspired by Shneiderman's Visual Information–Seeking Mantra, see Figure 2.2. First an overview is given with the geographic map which contains the groups for all 16 states of Germany, visualized by donut charts. This overview shows the distribution of the selected category data for each state.

Using various filters, the data displayed in the overview for the selected filters is updated and then can be viewed again. Selecting one of the groups selects all the participants that are in this group to show them in more detail.

The details are then visualized in pie charts for each participant and other charts like radar, scarf or scatter plots, depending on the selected category. With the selection of specific time waves or participants, the detailed view is showing information for these selected subjects. From the details, participants or time waves can be evaluated and set as new filters, to inspect other groups or participants. With the use of this workflow, the user can get an overview over the MoMo data and with some clicks and interactions on the system gain new insights and ideas which could be investigated further.
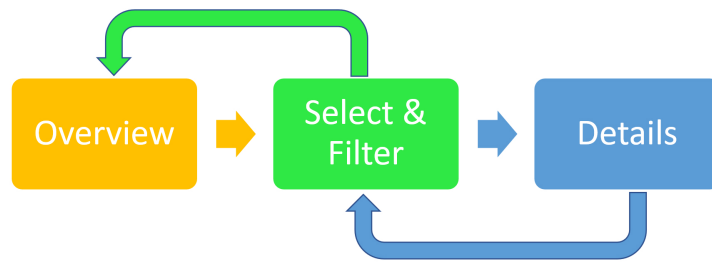
Figure 3.17: Workflow in VASD, first overview, then select and filter, then more detailed information for the selection.

4 Evaluation

To improve the VASD system and see how users interact with it a user study was done after finishing the major tasks of visualizing the provided MoMo data. The user feedback led to finding and fixing bugs that did not appear previously due to other work environments (different screen resolutions or web browsers used). There were also some tweaks made to the visual representation of data and with these tweaks the comprehensibility of the VASD system was improved.

4.1 Procedure

This user study was done in several online meetings. The 1–hour long meetings were scheduled with an online calendar and carried out in Webex [Cis] meetings. First the participants had to agree to a consent form that their usage of the VASD system would be recorded and analyzed. After this there was an around 10 to 15 minutes long introduction to the system, where every view and function as well as how to use it was explained. Then the users had time to try out the system as long as they wanted in an explorative approach. The VASD system was hosted on a small webserver on bwCloud [LE] and the participants shared their screen, which then was recorded. When participants finished their usage of the system, they filled in a small questionnaire about the usability and usefulness of the system. Questions about participants demographic characteristics were also asked, such as age, gender, employment status, research area, and familiarity with visualizations used in the system. Improvements to the system and other things that should be changed about the system could also be indicated in the questionnaire.

4.2 Participants

The user study had 7 participants, 3 were female and 4 male, from age 26 to 31 with a mean age of 28. Chosen participants were from the KIT, 6 Ph.D. students and 1 master student from different departments, four from the department of Health, two from Sports psychology and one from Sport and Space. Almost all participants use statistics in their daily work, see Figure 4.1. Other demographic questions were about the familiarity with different forms of visualization. All participants sometimes use pie charts in their work, five participants sometimes use scatter plots and two use them regularly in their work. The use of radar charts was not widely seen, only one participant uses it for their work, one participant outside their work and five participants heard about this type of visualization. The participants already were familiar with the visualized data before this study, some were domain experts. Different web browsers were used by the participants, Chrome [Goo] was used by 3 participants, Firefox [FC] also was used by 3 participants and 1 participant used Edge [Mic] for the study. Participants did not receive any compensation for doing the user study.
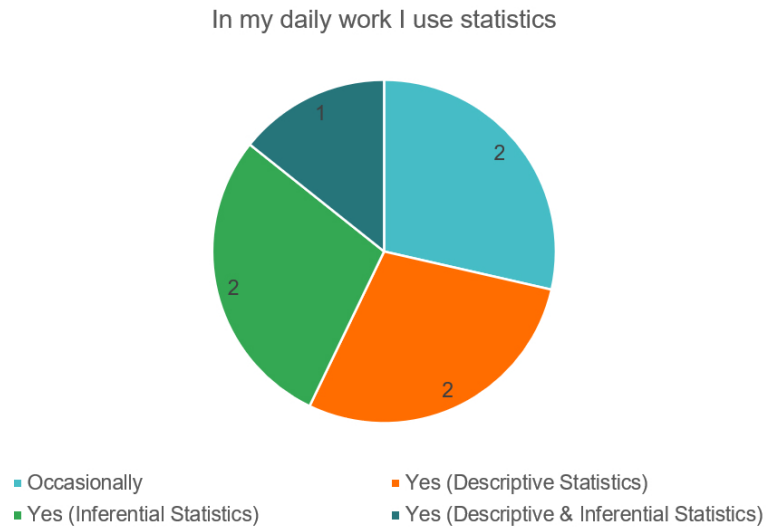
In my daily work I use statistics

- Occasionally
- Yes (Descriptive Statistics)
- Yes (Inferential Statistics)
- Yes (Descriptive & Inferential Statistics)

Figure 4.1: Answers to how often the participants of the user study use statistics in their daily work.

4.3 Questionnaire Results

To measure how easy it is to use the VASD system, the System Usability Scale (SUS) [Bro+96] was used. This is a ten–item questionnaire about the usability of a system, with five answer possibilities for each question, ranging from Strongly disagree to Strongly agree. The questions asked were:

1. I think that I would like to use this system frequently

2. I found the system unnecessarily complex

3. I thought the system was easy to use

4. I think that I would need the support of a technical person to be able to use this system

5. I found the various functions in this system were well integrated

6. I thought there was too much inconsistency in this system

7. I would imagine that most people would learn to use this system very quickly

8. I found the system very cumbersome to use

9. I felt very confident using the system

10. I needed to learn a lot of things before I could get going with this system

Each of the answer possibilities represents a point on the point scale, Strongly disagree is value 1 and Strongly agree value 5. Questions 1, 3, 5, 7 and 9 then calculate their value with their chosen point minus 1, for example Agree with a value of 4 becomes 4–1=3. Questions 2, 4, 6, 8, 10 get deducted from a start value of 5, for example Disagree with a value of 2 results in 5–2=3. After doing these calculations for all ten questions, the results are summed up and then multiplied with 2.5 to obtain the overall value of SUS [Bro+96], which can range from 0 to 100.

The results of this user study for all participants ranged from 65 to 85, with a mean of 73.21. This means the system is usable but needs some improvements to help new users with the handling of the system. Some additional questions in the questionnaire were related to getting to know and understand the MoMo data set using the different VASD visualizations, see Figure 4.2.
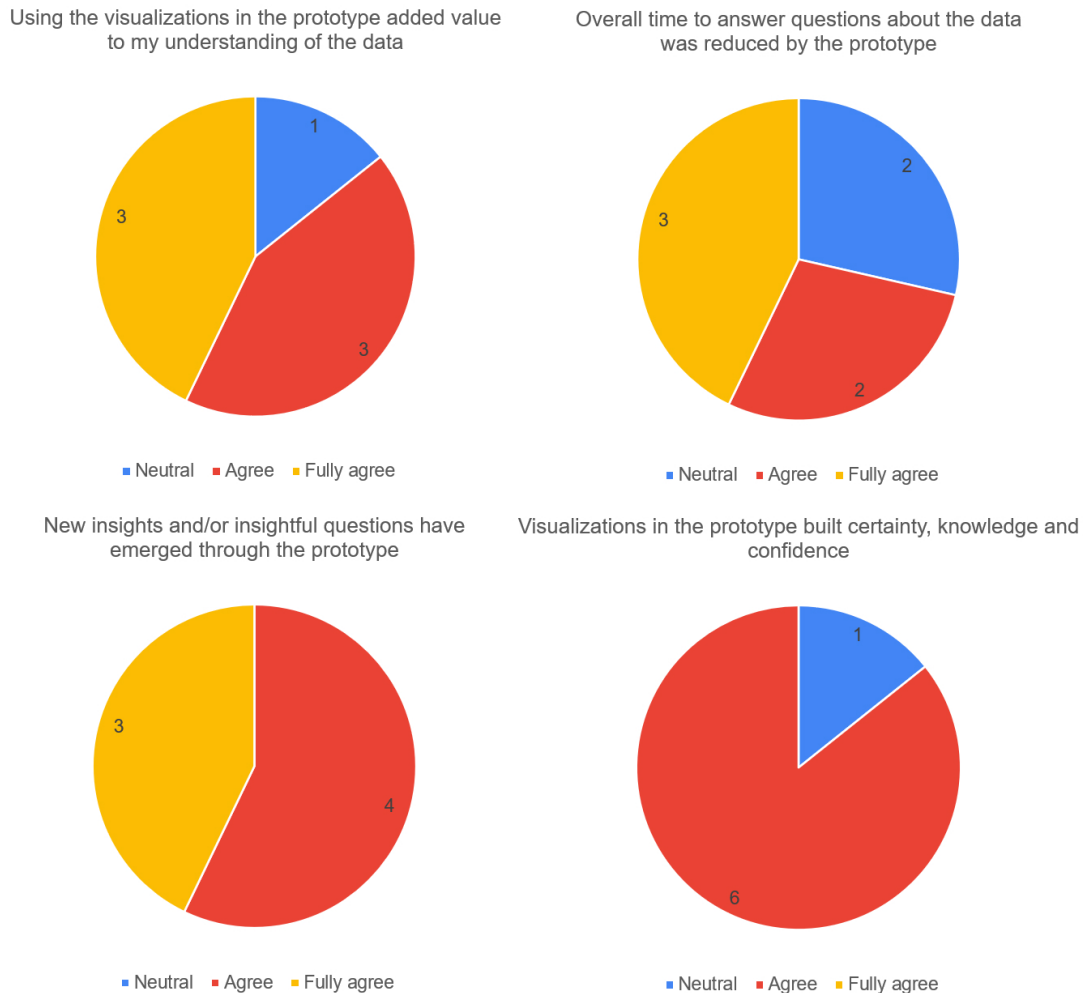
Figure 4.2: Questions about understanding the data through visualizations in the system.

These results show that the system improved understandability of the data set, reduced the time to find answers to considerations and helped in finding new insights and increase knowledge on the data set.

A suggestion made by participants to improve the use of VASD was to implement some form of figures exportation, to directly compare them with each other. With the current state of the system, a workaround would be opening the site two or more times, using the wanted filters for each of the sites and comparing the data by switching between the different pages. Another suggestion that was made from multiple participants was to allow direct comparisons between the categories, to display

them together and not separated to improve group or state comparisons. This could be a future feature of VASD, again with the current work around of opening the site multiple times, switching between them or comparing side by side.

For the mean values displayed on the scatter plot, there were suggestions to change this type of visualization to a box plot, to display the mean values in addition with the distribution of the data. The ability to change the X–axis and Y–axis data categories into other categories was also requested, to convert them into more appropriate categories for one's own analysis.

One suggestion that was implemented after user feedback was to add the weekday to the date shown on the Accelerometer scarf plot, to more easily tell the user on which day specific activities happened. With this addition, the comparison between weekends and weekdays was simplified.

Another suggestion was to change the color scale on the scarf plot, to more easily differentiate between Sedentary and Light activities. And a suggestion was made to display sources and explanations for the data, for example where physical fitness based data comes from or on what the percentiles are based.

4.4 Insights and Exploration Results

Participants of the user study tried all available categories and visualizations. The comparisons that were made showed good results and gave new insights and ideas about the underlying data. One finding for category BMI was that some participants had BMI values in time waves they did not participate in, this finding is described in more detail in Section 5.1.3. Another often used comparison was the Fitness comparison, described in Section 5.2.1. Also, a comparison between low and high social–economic status for the BMI was done as described in Section 5.2.1. An Analysis of the surroundings of participants was done as described in Section 5.2.2.

One interesting finding about the Accelerometer data was seen for a participant of the MoMo study, who had some form of activity that lasted one hour and was demanding. It was suggested that this was a soccer game on a Sunday, without knowing if the date seen was really this weekday, because the weekday was not displayed at this time. After checking the specific day, it really was a Sunday and could fit the described activity of playing soccer. A similar comparison of accelerometer data was done in Section 5.2.3.

Fixes that Resulted from the User Study Analysis

Quick interactions with the VASD system led to different errors when the data was not loaded fast enough. Some adjustments were made to compensate for this type of interaction speed or slower internet connections. There were also some problems with different resolutions for participants of the study, where elements were shown beneath each other when they should be beside each other. This was fixed by some smaller adjustments aligning the different views together. Errors that threw undefined data errors and resulted from interactions with the system were also fixed.

5 Analysis and Use Cases

In this section, various lessons learned from using VASD are explained and three use cases of the system analyzed.

5.1 Insights Found while Trying Out the System

There were some issues about the data set, which appeared while implementing new functions into the system. If the data in the visualizations did not look right and it was not a system related implementation problem, discussions were held with domain experts of KIT to understand why this was the case. Some of these issues resulted in changes to the processing of the data.

5.1.1 Alter_T5 Error Fix

After implementing the displaying of ages for different time waves of each participant, while looking at single participants there was an anomaly seen for time wave T5. For these participants there were negative values shown, which seemed like there was a typo while entering the age, for example in T4 the age was 5, for T5 –6 and for T6 7, see Figure 5.1. Investigating more deeply all the data for Alter_T5, there were many negative values or values above 900, which showed that the entered ages were wrong. After feedback to the KIT, they checked for this and found a syntax error in their calculation of the age. Through the visualization of the data this syntax error was found, which shows that visualization is useful for analyzing data sets.

ID of participant: 722530
Age of participant: T4: 5, T5: -6, T6: 7
Gender of participant: weiblich
Timeframe of study: T4 (2018-2020), T5 (2020-2021),
T6 (2021-2021)
State of participant: Schleswig-Holstein

Figure 5.1: Alter_T5 is shown incorrectly as –6.

5.1.2 Finding Irregularities in the Data Set

While doing data cleansing on the data set there was an error for the data in year of study „Untj_T3" and month of study „Untm_T3." The year and month fields were switched for this time wave, which then could be solved afterwards. This was only a problem for the separately created MoMo data set for this thesis.

5.1.3 BMI for More Time Waves than Participated

Looking through category BMI KH pie charts for the time waves T1–T6 of participants in a group, there was one issue that caught the eye. For time wave T1 BMI KH data was available but from the tool–tip shown there was no participation in this time wave from the participant, see Figure 5.2.

After asking the domain experts at KIT about this, the answer was that MoMo longitudinal participants of time wave T2 had information requested for time wave T1 and this included BMI, to increase the amount of information about the participants. Therefore, this data is available, but the participants only participated in time wave T2 and not in time wave T1. There are also participants from whom there is simply no information for T2 because they only filled out the activity questionnaire but did not participate in any other part of the survey, for example the motoric skills tests. This is individual for each participant.
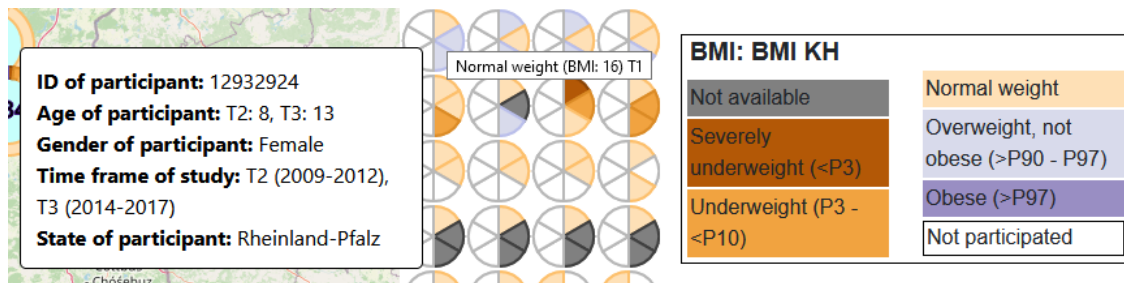

Figure 5.2: BMI and BMI KH for T1 is available, but no participation in T1.

5.1.4 BMI KH and BMI IOTF Wrongly Classified

Checking out different categories for the time waves T4–T6 to see changes in fitness and health during the Corona virus pandemic, some irregularities appeared in category BMI KH. Some participants had a change in their division of the BMI for time waves T4 and T6. Hovering over the tool–tip for more information, these participants had the same BMI in both time waves, but the division changed significantly. Participant 810816 even had a higher BMI in T6 (33.4) than in T4 (29.9) but was classified into Normal weight for T6 and Obese for T4, see Figure 5.3. This was verified by the domain experts from the KIT to be a mistake and was fixed with a new BMI KH data set. With the help of visualization and the use of tool–tips this error was found and underlines the usefulness of visualizations.
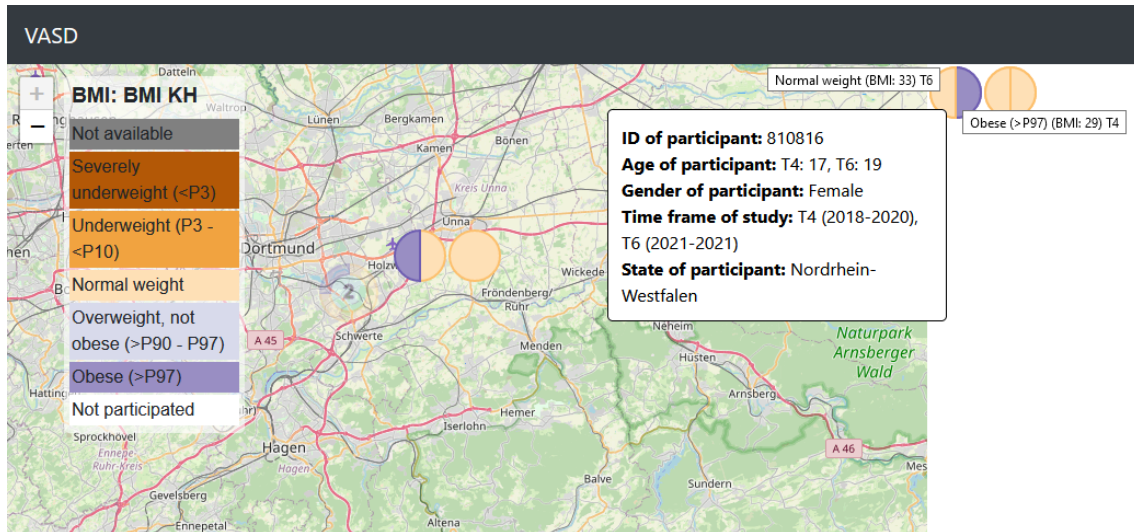
Figure 5.3: Participant 810816 has higher BMI in T6 than in T4, but is classified better.

5.2 Use Cases

Three use cases are detailed in this section, covering all categories available in VASD (Accelerometer, Fitness, BMI and Questionnaire). The first use case involves fitness and health comparisons between different groups of participants, the second use case again involves comparisons between different groups of participants and their environment, and the third use case involves more detailed analyses of a single or multiple participants.

5.2.1 Fitness and Health Comparison

The comparison of participants for their fitness and health is possible with the categories Fitness and BMI. Ideas to analyze this would be to ask: Do kids who are in sports clubs have better fitness and health? Do kids from socially higher families have healthier BMI scores? For these questions, the filters sports club member and social status can be used. To compare yes and no more easily with each other opening the VASD system on a second tab or doing screenshots of the visualizations is possible. Opening the system in two tabs would be more practical because then the interactability is still present. In Figure 5.4 the overview of the category Fitness is seen for sports club members on the left and participants who are not sports club members on the right side. Selected is the state Sachsen–Anhalt, to show mean values for the different fitness tests and the relation between Total fitness and BMI.

Directly visible is that the groups differ significantly between the two options, sports club members are much more often in the divisions 50–85 and 85–100 percentile, while non–members are mostly in 15–30 and 30–50 percentile. This shows that participants who are members of a sports club are fitter than non–members.
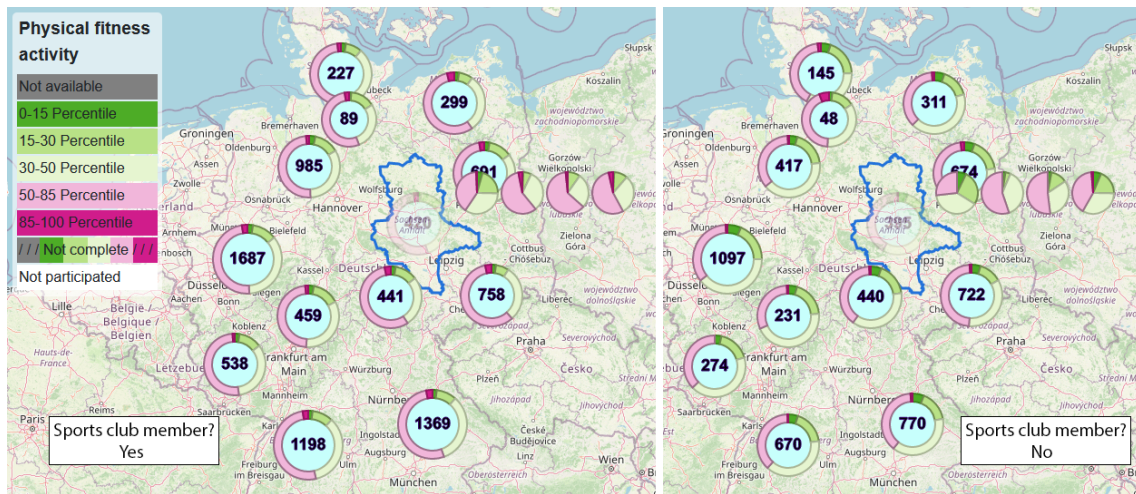
Figure 5.4: Total fitness overview for sports club members on the left and non–members on the right. Sports club members more likely to be in 85–100 percentile range.

Comparing the mean values for both options, the sports club members again have better overall scores, see Figure 5.5. For the sports club members, the total fitness score is never below 45 for any of the time waves and for the non–members it is below 40 for time wave T1. Regarding the highest mean scores, the sports club members have a max value of 57, while the non–members never reach above 52. Interestingly the relation between Total fitness and BMI is not that different for both options. As previously seen, the actual total fitness scores are higher for sports club members and the BMI is marginally better.
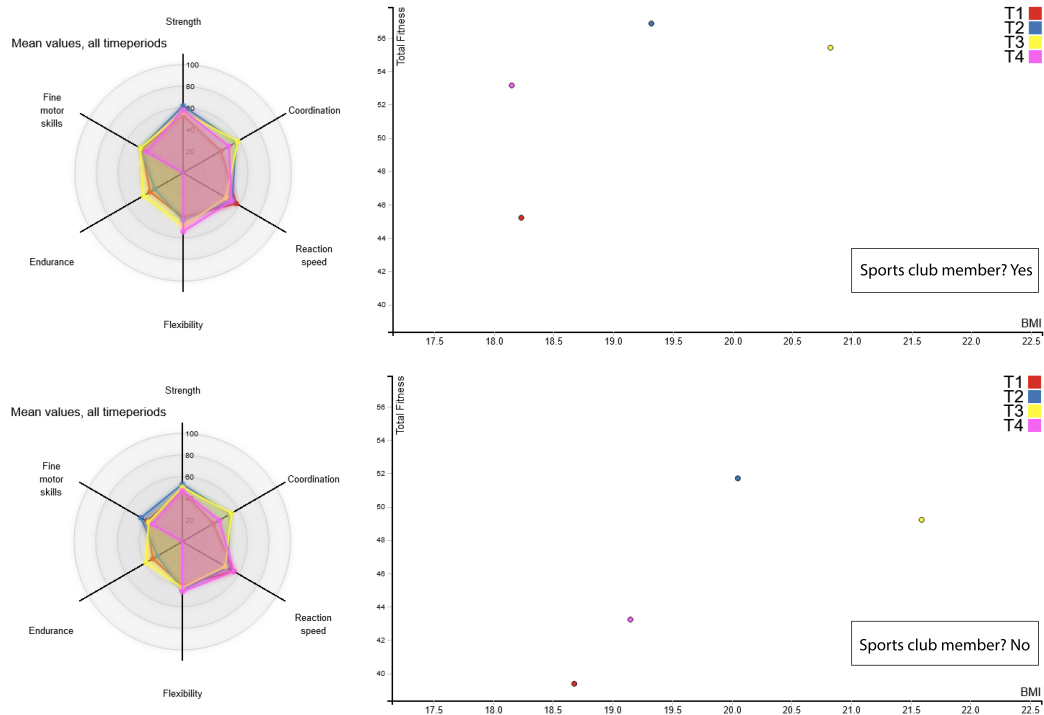
Figure 5.5: Sports club members mean values on top and non–members mean values below for Sachsen–Anhalt. Sports club members have higher total fitness scores.

In the overview for category BMI a similar pattern can be observed as in category Fitness, where sports club members are more often in division Normal weight, while non–members are more often in the divisions overweight and obese, see Figure 5.6. This indicates that sports club members are more often fitter and healthier than non–members, which should be no surprise because in a sports club you are at least active during the club training. Children which do not do sports are also more likely to not be a member of a sports club, resulting in the higher amount of overweight and obese participants in the non–member part.
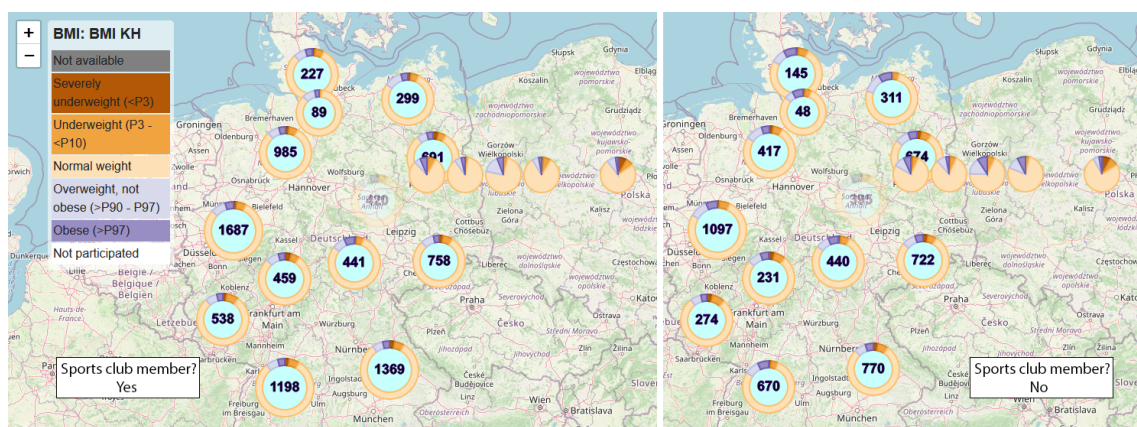

Figure 5.6: BMI KH distributions for sports club members (left) and non–members (right). Non–members more likely to be overweight or obese.

Another comparison for the category BMI with the filter social status is done, see Figure 5.7. This comparison between participants with high or low social status shows a difference in the number of overweight or obese participants. Participants from lower social status families were more likely to be in one of these divisions compared with higher social status participants. Although fewer participants were classified as low social status, more of these participants were overweight or obese in one or more of the time waves than the higher number of high social status participants. In addition, socially weaker longitudinal participants were more likely to remain in the division overweight or obese across multiple time waves compared with socially stronger longitudinal participants.
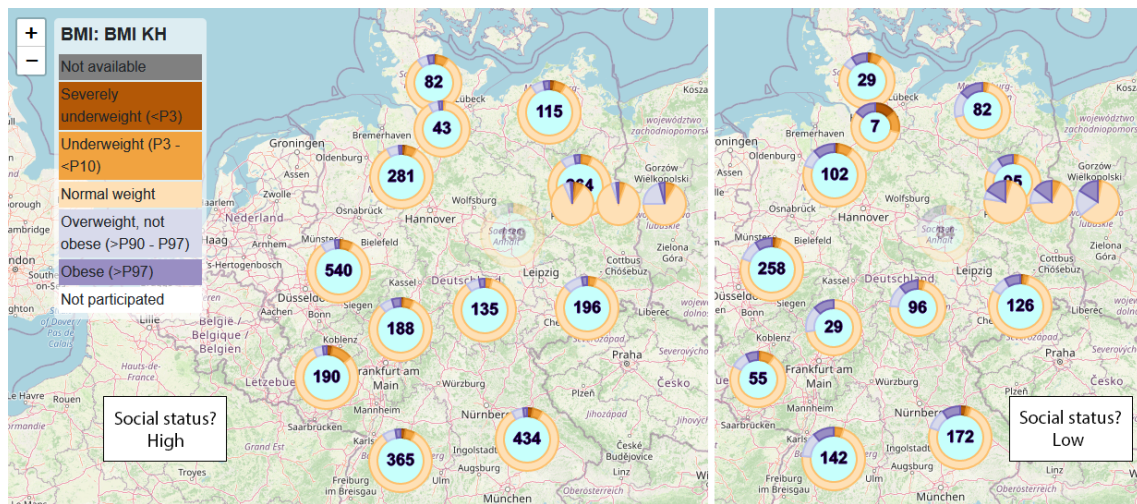
Figure 5.7: Distribution of BMI KH for high social status (left) and low social status (right). Low social status participants are more likely overweight and obese.

As a last impression for all participants without any filter applied the state Sachsen–Anhalt is again inspected and this time only time wave T2 selected. The result from this is seen in Figure 5.8 and it shows that there is a relation between healthy BMI scores and good Total fitness results. Participants with higher BMIs (above 30) are more likely to perform worse in these fitness disciplines, still the results below this BMI value differ significantly.

Figure 5.8: Time wave T2 mean values and relation between Total fitness and BMI for all T2 participants. The higher the BMI, the lower the total fitness scores are.

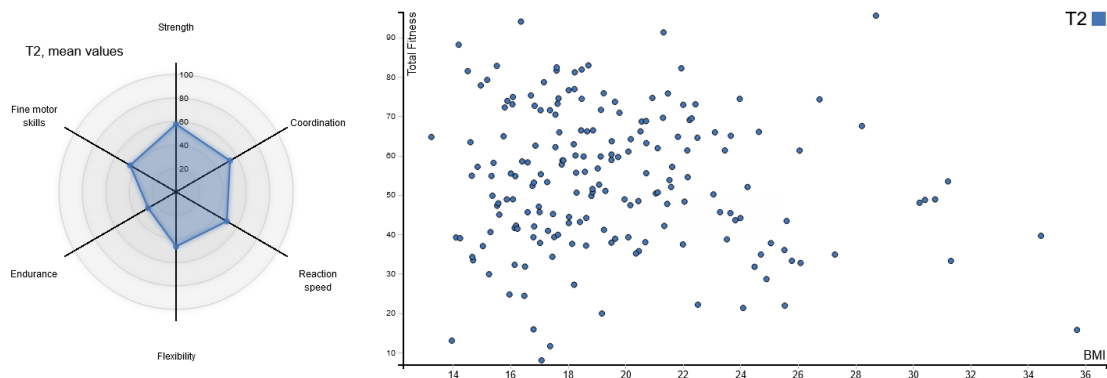### 5.2.2 Surroundings of Participants Analyzed

Differences in participants surroundings can potentially influence sporting activity, thus a comparison of major cities with rural towns could be a useful insight. For this comparison the category Questionnaire: School route is selected with the filter City applied for Major city (left) and Rural town (right), the overview can be seen in Figure 5.9. On this overview, the state Sachsen is selected to see relations between the Total Fitness and the selected subcategory School route for all time waves. The way to school in the major cities is much more often done on foot or by bicycle than in rural towns, where moving by bus or train and by car or moped is more prominent. This observation shows that the usage of public transportation and personal vehicles is higher in rural towns, which could be an indication that children from these towns must move into other cities to get to their schools.

Another observation that can be seen for both types of cities, especially rural towns is that the relations between all the divisions are not equal across Germany. In northern states much more often the car or moped was used to get to school compared to the south. Additionally, in the south more often a bus or train was used to get to school, which could be a result of better public transport infrastructure, great differences can be seen here in Baden–Württemberg compared to Hamburg and Schleswig–Holstein.
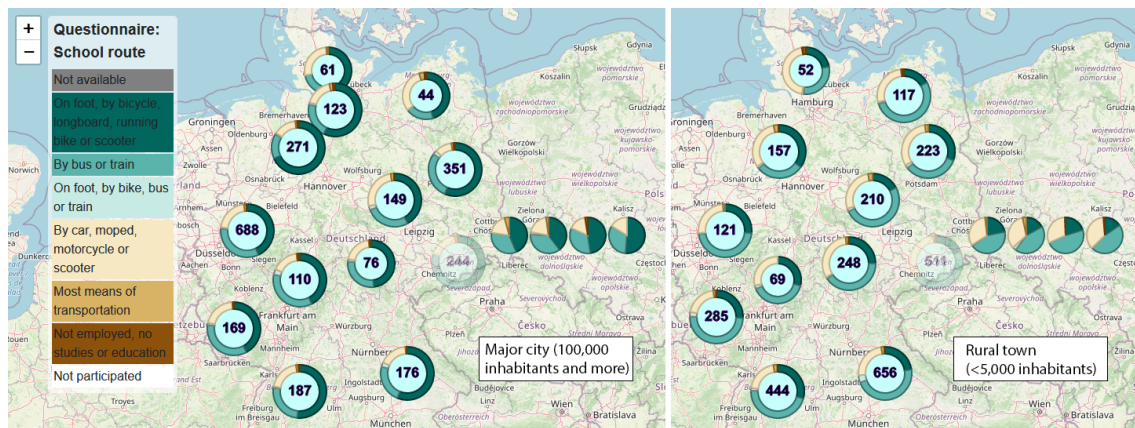
Figure 5.9: Comparison of major cities (left) and rural towns (right) way to get to school. Participants in major cities more often go to school on foot or by bicycle.

The pie charts for each participant in the selected group in rural towns visualize for longitudinal participants how their way to school changed over time, see Figure 5.10. Some children at a young age went to elementary school by walking there on foot or by bicycle, later they went to high school by bus or train and when old enough they got there with a car or moped.

Figure 5.10: Change over time how children go to school in a rural town (<5000 inhabitants), first on foot or bicycle, then by bus or train and when old enough by car or moped.

Comparing the relations between total fitness and the divisions of the subcategory School route for different time waves in Sachsen does not yield huge differences between participants fro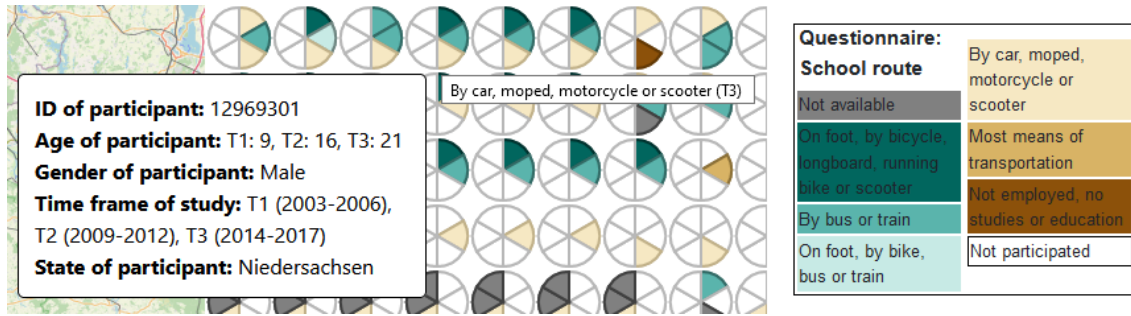m major cities or rural towns. In some time waves the major city participants far exceeded the rural town participants and in other time waves this was the other way around, see Figure 5.11. It should be noted that the number of participants in Sachsen from rural towns is more than double the number of participants from major cities, which could distort the results.
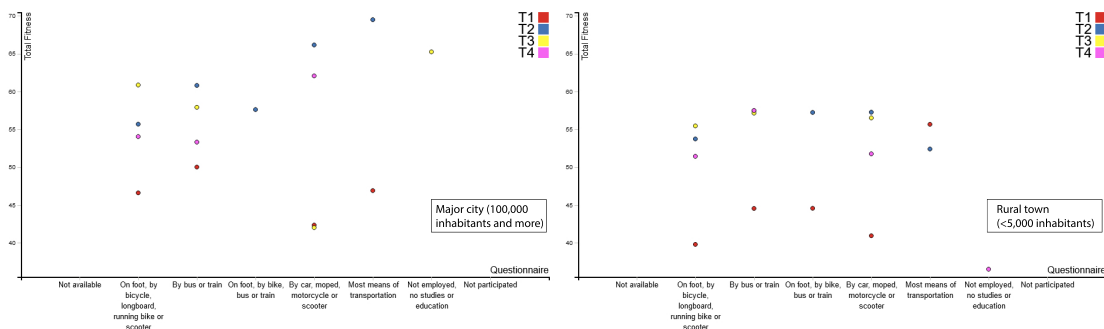
Figure 5.11: Relation between Total fitness and subcategory School route does not show huge differences between city sizes.

One last comparison for public infrastructure is done with the subcategory Bike lanes nearby, which could tell the reasons why in some areas seen in Figure 5.9 more children travel to school on foot or bike than in others. This seems to be a case in some states, mostly in the north, see Figure 5.12, where more children chose Many bike lanes nearby in major cities than in rural towns.
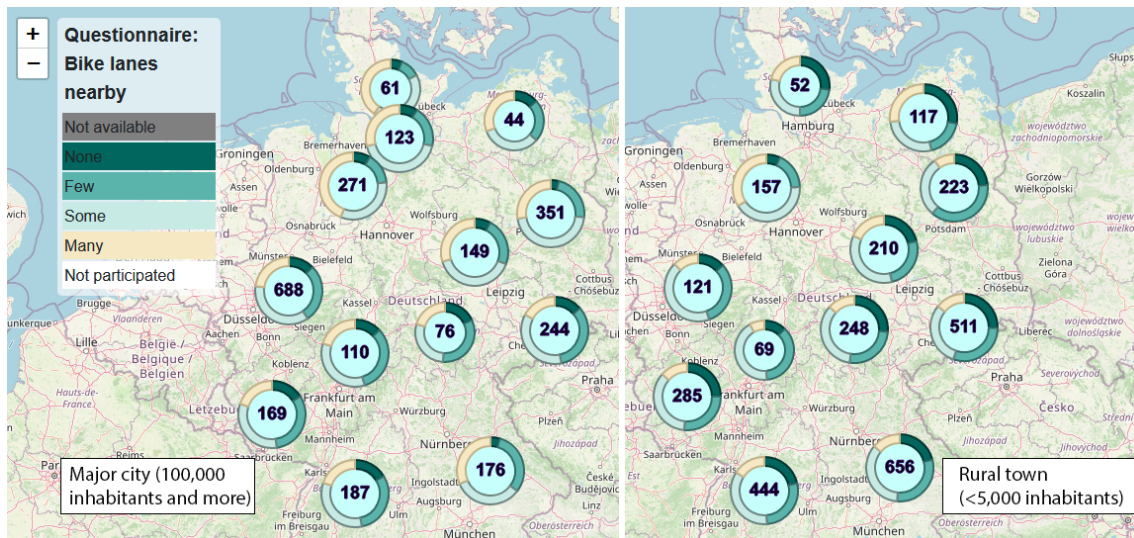
Figure 5.12: Comparison of Bike lanes nearby between major cities (left) and rural towns (right). Better developed in major cities.

In conclusion, the different visualizations showed that differences can be seen in major cities compared to rural towns, which can lead to an increase or decrease of specific means of transportation depending on the public infrastructure.

5.2.3 Analyze One or Multiple Participants

Analyzing one or more participants in more detail is possible in VASD. For this evaluation the category Accelerometer is used, a date range from 01/12/2015–01/17/2015 chosen, the filter to look for every participant under or age 15 and all time waves T1–T6 are selected. On selection of these restrictions on the data set, only five participants are left to select accelerometer data from, which are all located in Hessen. Selecting these participants returns the accelerometer data for this date range, see Figure 5.13.
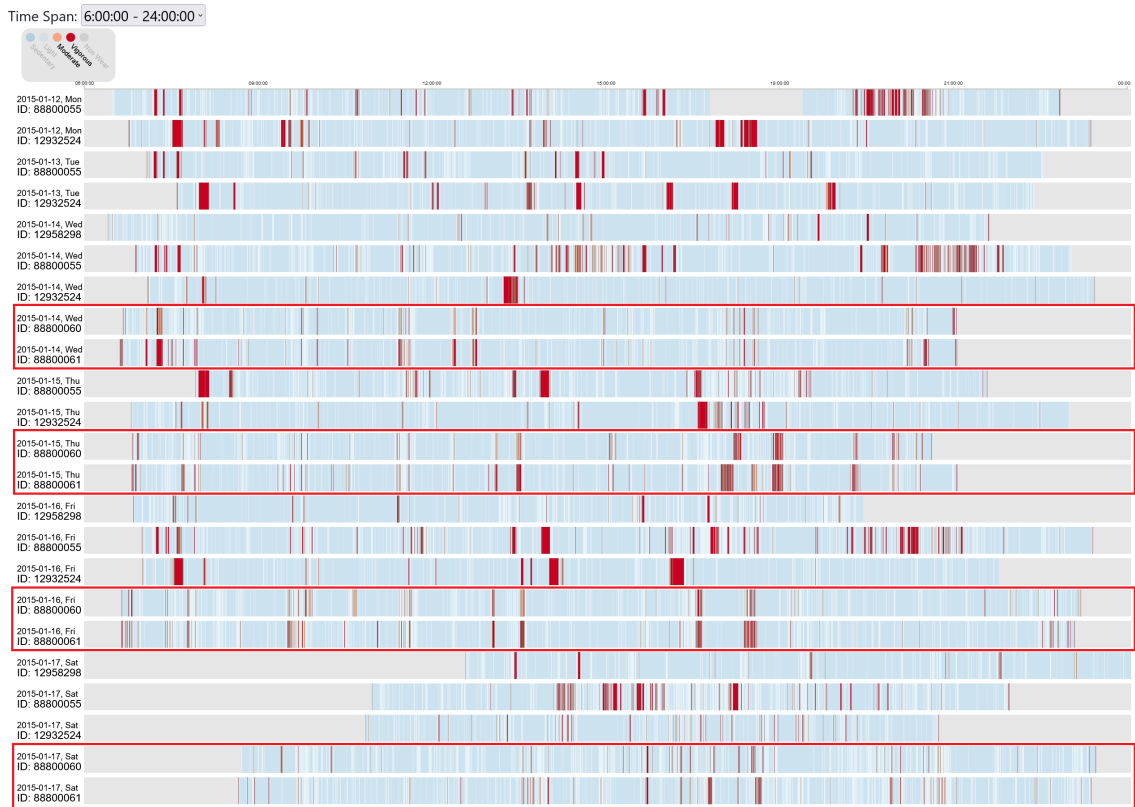
Figure 5.13: Accelerometer data for the five selected participants. Participants 88800060 and 88800061 stand out with similar activities.

Looking for Moderate or Vigorous activities in the data from 6 to 24 o'clock brings up two participants with the IDs 88800060 and 88800061, which look nearly the same regarding their daily activities highlighted red in Figure 5.13.

Upon selection of these two participants with the filter Participant IDs and again looking at the accelerometer data, it looks like these two participants are related to each other because their data almost matches exactly when they do strenuous activities and when they start or end their day doing activities, see Figure 5.14. Looking at the tool–tips for both participants on the pie charts, shows that both only participated in time wave T3 of MoMo, are both male, aged 11 and have a near identical ID, which in total could mean they are twins.
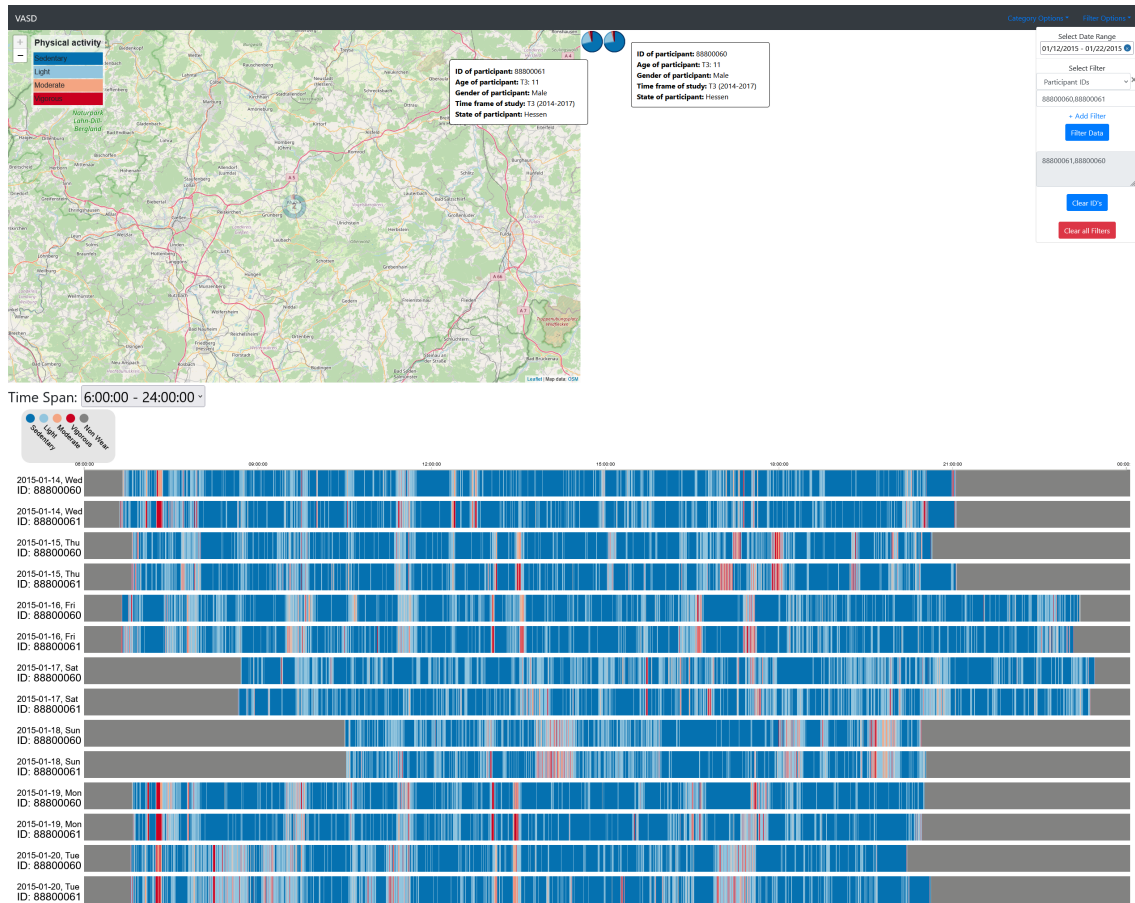
Figure 5.14: Accelerometer data only for participants 88800060 and 88800061 displayed. Wearing times and type of activity match almost exactly.

Switching to the category Fitness, the two participants are both categorized into the 15–30 percentile range, with both having similar results for each test, except participant 88800060 being much better on reaction speed but has a lower total fitness score, see Figure 5.15.

In category BMI KH, the participants were checked for their BMI and how many days they are active in a week doing sports. With a BMI of 21.16 participant 88800060 has a BMI KH of normal weight, while participant 88800061 with a BMI of 24.16 is in division overweight, but not obese. Both are active during a week with participant 88800060 being active 5.5 days and participant 88800061 with 6 days of activity in a week.
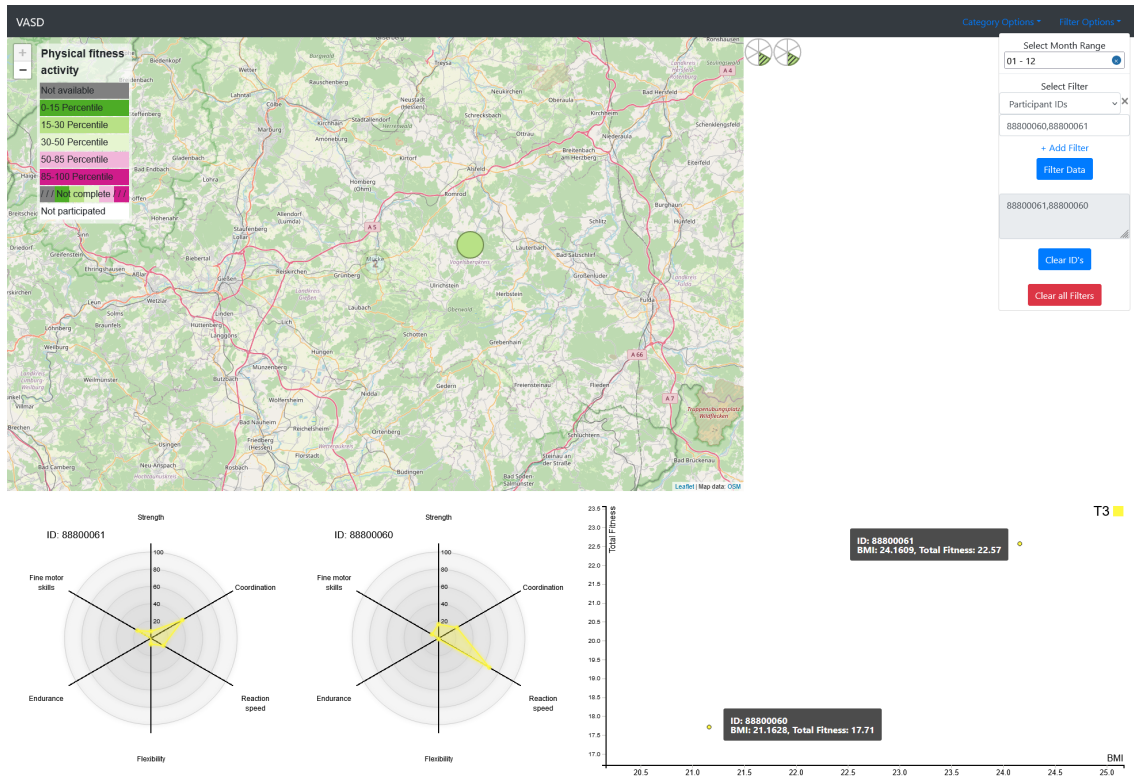
Figure 5.15: Fitness scores for both participants 88800060 and 88800061. Radar charts show no huge difference except for participant 88800061 being better at coordination, while 88800060 is better at reaction speed.

For category Questionnaire (School route) both participants selected that they get to school by bus or train, maybe matching with the accelerometer data seen in Figure 5.14, being moderate or vigorous for the times 07:05:00–07:06:30, 07:14:00–07:19:15, around 13 o'clock and 13:12:30–13:24:30, which could indicate running to a bus or train to catch it or running to school from a bus or train arriving in time, see Figure 5.16. The other questions of the questionnaire were also answered with nearly matching results, which could strengthen the consideration that these two participants are twins.
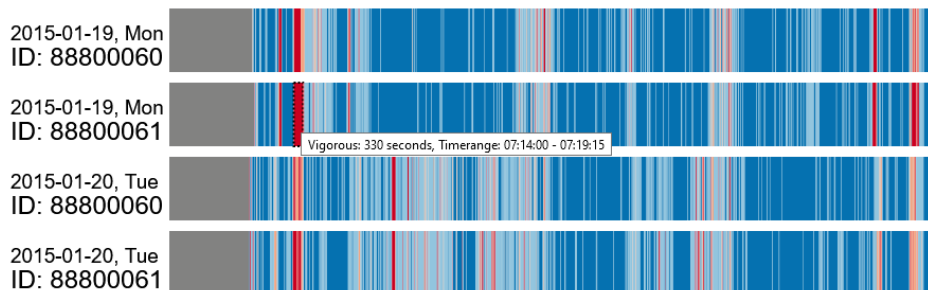
Figure 5.16: Moderate and vigorous activity indicating running to or from a bus or train to get to school for participants 88800060 and 88800061.

6 Conclusion, Limitations and Future Work

This section sums up the work done to implement and realize VASD, shows limitations of the system and presents future work that can be done to improve it further.

6.1 Conclusion

In this work the VASD system was presented which visualizes the MoMo data set of children and teenagers to support users with their exploratory research on the data. The related work was first presented to show how other projects realized various visualizations for the use in a fitness system and how different approaches to visualization are seen from the user perspective. After introducing related work, the processing of the provided MoMo data was described and how these different data sources were combined to introduce a way of analyzing them together. Then a general system overview was given, which explained the four categories with their specific approaches to visualize data and all other functionalities of the system that could be used to filter data, a combination of time selection and multiple filters. An evaluation was done on the system with a small user study, which showed good results for the usability of the system, with some adjustments needed and that VASD can be used to explore the data. Three use cases were presented which showed different approaches to visualize MoMo data and how to use the system to gain new insights and impressions on this data while analyzing it.

With the use of VASD the quality of the MoMo data set was improved, because the use of the visualizations brought up issues that were then fixed. Analyzing accelerometer data from MoMo participants with VASD to find patterns of higher activity is a good way to learn about children's and teenagers behavior. The other visualizations in the system can be used to analyze groups of participants to find differences and trends between states or other distinguishing characteristics. Using the filters and time selections offers the chance to analyze groups of participants by age, gender, their surroundings, and other filters over different time waves at selected time periods of the MoMo survey. The system also allows for the analysis of longitudinal participants over multiple time waves to recognize changes in their environmental surroundings or behavior. In summary, the VASD system is a useful tool to visualize the MoMo data set for quick overviews from different perspectives and offers on demand more in depth analysis for specific questions.

6.2 Limitations of the VASD System

For category Accelerometer only data of ten participants can be shown simultaneously because it takes a long time to load and process the data. For faster interaction with this category the use of filters and the date range selection is recommended. Interactability is limited in the more detailed charts, only hovering to get tool–tips with extra information is available and for example time waves selection on the scatter plot legend is not possible.

It could be irritating to filter for age under 15 and receive persons who are older in a time wave, because the filter process looks for every selected time wave and if there is a match it gets added to the visualized data. This can be limited by manually selecting the wanted time wave under the category Options. Currently, it is not possible to add more than one of the same filters or use multiple ranges in one filter like age range, for example filter after age 5–10 or 15–20, or filter for ages 6, 8, 10, 12 only.

Furthermore, analyzing across different states is not possible, which means there can not be direct comparisons between Bayern and Berlin or other states. Comparisons between participant groups are also not possible, but a small workaround could help. The workaround would be to open multiple browser windows with the VASD website and select the filters and settings separately or use only one site and do screenshots to compare them later, which would limit further detailed analysis.

Privacy Reduces Analyzability

Because only the state is shown for each participant in the groups on the geographic map, analyzing trends in surroundings is harder to do. The surroundings can only be analyzed by looking at the questionnaire the participants filled out, sometimes in support with their parents. Walkways to school or playgrounds in the neighborhood can not be seen by looking directly at the map. Because all the participants of a state are clustered together there can not be an analysis if some region in a city has more healthy children, as a result of their surroundings or if there are other reasons, why they are healthier. Also, the analysis of friend groups, which go to school each day, or meet each other to play together is harder to do.

6.3 Future Work

Improvements that could be made to the VASD system are faster loading and serving of the different views available, especially the long loading times for the scarf plot in category Accelerometer. One thing that could be done to improve the responsiveness would be to save calculation scores into the database, because currently for example the total fitness score gets recalculated each time it is needed. The accelerometer data could also be summed up to reduce the loading process that occurs before the data is transmitted to the system's visualization.

Another improvement for the system would be to analyze the visualized data across multiple categories simultaneously, currently only data for one category is always displayed. This also applies to comparisons between participant groups, for example comparison of participants from different states or a comparison of female and male participants.

The sorting of the pie charts could be changed, to fit another preference of their allocation. As suggested in Section 4.3, a box plot could be implemented instead of the scatter plot, to show minimum and maximum values in addition to the mean values. If the scatter plot remains in the system, the data categories for both axes should be made selectable and the minimum and maximum values of the data set should be used on the axes, to improve comparability.

To allow for a public release of the VASD system the Flask API must be exchanged, because it is not intended for public infrastructure.

Adding more MoMo time waves would make it necessary to update the selectable time waves of VASD. Participants data could receive additional detailed fitness activities, for example, heart rate tracking or ergometer tests to allow an analysis of this sort of fitness and health data. With the use of other algorithms or cut–points, accelerometer data for participants older than 17 years in a time wave could be used. This would support the analysis of accelerometer data for longitudinal participants over multiple time waves. The use of machine learning could help with highlighting specific circumstances in the data set, finding new insights, or improving data representation and other possibilities.

Bibliography

[ABI19]   R. Aravind, T. Blascheck, P. Isenberg. "A survey on sleep visualizations for fitness trackers". In: *Posters of the European Conference on Visualization (EuroVis)*. 2019 (cit. on p. 20).

[Aga]   V. Agafonkin. *Leaflet*. URL: https://leafletjs.com/ (cit. on p. 30).

[AGPR06]   N. Ali-Hasan, D. Gavales, A. Peterson, M. Raw. "Fitster: social fitness information visualizer". In: *CHI'06 Extended Abstracts on Human Factors in Computing Systems*. 2006, pp. 1795–1800 (cit. on pp. 25, 26).

[Apa]   Apache. *Apache Parquet*. URL: https://parquet.apache.org/ (cit. on p. 30).

[BBB+18]   T. Blascheck, L. Besançon, A. Bezerianos, B. Lee, P. Isenberg. "Glanceable visualization: Studies of data comparison performance on smartwatches". In: *IEEE transactions on visualization and computer graphics* 25.1 (2018), pp. 630–640 (cit. on pp. 19, 20).

[BDHO]   M. Bostock, J. Davies, J. Heer, V. Ogievetsky. *D3.js*. URL: https://d3js.org/ (cit. on p. 30).

[BHU]   C. Brewer, M. Harrower, T. P. S. University. *Color Brewer 2*. URL: https://colorbrewer2.org/ (cit. on p. 38).

[BMA+20]   A. Burchartz, K. Manz, B. Anedda, C. Niessner, D. Oriwol, S. C. Schmidt, A. Woll. "Measurement of physical activity and sedentary behavior by accelerometry among a nationwide sample from the KiGGS and MoMo study: study protocol". In: *JMIR Research Protocols* 9.7 (2020), e14370 (cit. on pp. 15, 30, 38).

[Bro+96]   J. Brooke et al. "SUS-A quick and dirty usability scale". In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7 (cit. on p. 52).

[BSBE17]   T. Blascheck, M. Schweizer, F. Beck, T. Ertl. "Visual comparison of eye movement patterns". In: *Computer Graphics Forum*. Vol. 36. 3. Wiley Online Library. 2017, pp. 87–97 (cit. on p. 21).

[Buna]   S. Ä. des Bundes und der Länder. *Data set for Regionalatlas Deutschland*. URL: https://www.statistikportal.de/de (cit. on p. 27).

[Bunb]   S. Ä. des Bundes und der Länder. *Regionalatlas Deutschland*. URL: https://regionalatlas.statistikportal.de/ (cit. on pp. 26, 27).

[Cis]   Cisco. *Cisco Webex*. URL: https://www.webex.com/ (cit. on p. 51).

[CLMB11]   L. Choi, Z. Liu, C. E. Matthews, M. S. Buchowski. "Validation of accelerometer wear and nonwear time classification algorithm". In: *Medicine and science in sports and exercise* 43.2 (2011), p. 357 (cit. on p. 31).

[CMS99]   S. K. Card, J. D. Mackinlay, B. Shneiderman. "Information visualization". In: *Readings in information visualization: using vision to think* (1999), pp. 1–34 (cit. on pp. 16, 17).

[Com]   A. for Computing Machinery. *Association for Computing Machinery*. URL: https://acm.org/ (cit. on p. 26).

[DF]   R. Dahl, O. Foundation. *Node.js*. URL: https://nodejs.org/ (cit. on p. 29).

[ECG+08]   K. R. Evenson, D. J. Catellier, K. Gill, K. S. Ondrak, R. G. McMurray. "Calibration of two objective measures of physical activity for children". In: *Journal of sports sciences* 26.14 (2008), pp. 1557–1565 (cit. on p. 31).

[EE]   I. of Electrical, E. Engineers. *Institute of Electrical and Electronics Engineers*. URL: https://www.ieee.org/ (cit. on p. 25).

[EN]   B. Eich, Netscape. *Javascript*. URL: https://www.ecma-international.org/publications-and-standards/standards/ecma-262/ (cit. on p. 29).

[FC]   M. Foundation, M. Corporation. *Mozilla Firefox*. URL: https://www.mozilla.org/en-US/firefox/new/ (cit. on p. 51).

[FMKK21]   M. Franke, H. Martin, S. Koch, K. Kurzhals. "Visual Analysis of Spatio-temporal Phenomena with 1D Projections". In: *Eurovis 2021-23rd EG Conference on Visualization*. 2021 (cit. on p. 23).

[Goo]   Google. *Google Chrome*. URL: https://www.google.com/chrome/ (cit. on p. 51).

[Gov]   Govdata. *GovData DL-DE->BY-2.0*. URL: https://www.govdata.de/dl-de/by-2-0 (cit. on p. 27).

[Hip]   D. R. Hipp. *SQLite*. URL: https://sqlite.org/ (cit. on p. 30).

[HNA+11]   K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, J. Van de Weijer. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011 (cit. on p. 21).

[HS12]   J. Heer, B. Shneiderman. "Interactive dynamics for visual analysis". In: *Communications of the ACM* 55.4 (2012), pp. 45–54 (cit. on pp. 16, 23).

[Ins]   R. K. Institute. *Robert Koch Institute*. URL: https://www.rki.de/EN/Home/homepage_node.html (cit. on p. 15).

[JRK+17]   M. John, C. Richter, S. Koch, A. Kuczera, T. Ertl. "Interactive Visual Exploration of the Regesta Imperii." In: *DH*. 2017 (cit. on p. 24).

[KFBI21]   C. Kinkeldey, J.-D. Fekete, T. Blascheck, P. Isenberg. "BitConduite: Exploratory Visual Analysis of Entity Activity on the Bitcoin Network". In: *IEEE Computer Graphics and Applications* (2021) (cit. on p. 21).

[KHD+10]   D. A. Keim, M. C. Hao, U. Dayal, H. Janetzko, P. Bak. "Generalized scatter plots". In: *Information Visualization* 9.4 (2010), pp. 301–311 (cit. on p. 22).

[KHW14]   K. Kurzhals, F. Heimerl, D. Weiskopf. "ISeeCube: Visual analysis of gaze data for video". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. 2014, pp. 43–50 (cit. on pp. 21, 22).

[KTW+13]   R. Krüger, D. Thom, M. Wörner, H. Bosch, T. Ertl. "TrajectoryLenses–A Set-based Filtering and Exploration Technique for Long-term Trajectory Data". In: *Computer Graphics Forum*. Vol. 32. 3pt4. Wiley Online Library. 2013, pp. 451–460 (cit. on pp. 23, 24).

[LE]   O. Landesprojekt "bwCloud SCOPE - Science, Education". *bwCloud*. URL: https://www.bw-cloud.org/ (cit. on p. 51).

[MBRC15]   J. Munoz, S. B. i Badia, E. Rubio, M. Cameirao. "Visualization of multivariate physiological data for cardiorespiratory fitness assessment through ECG (R-peak) analysis". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2015, pp. 390–393 (cit. on pp. 20, 21, 24, 25).

[Mic]   Microsoft. *Microsoft Edge*. URL: https://www.microsoft.com/edge (cit. on p. 51).

[NUO+20]   C. Niessner, T. Utesch, D. Oriwol, A. Hanssen-Doose, S. C. Schmidt, A. Woll, K. Bös, A. Worth. "Representative percentile curves of physical fitness from early childhood to early adulthood: the MoMo study". In: *Frontiers in public health* 8 (2020), p. 458 (cit. on pp. 32, 38).

[OW00]   T. J. Overbye, J. D. Weber. "Visualization of power system data". In: *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*. IEEE. 2000, 7–pp (cit. on p. 19).

[Pal]   Pallets. *Flask*. URL: https://flask.palletsprojects.com/en/2.0.x/ (cit. on p. 29).

[PSBS19]   C. S. Peterson, J. A. Saddler, T. Blascheck, B. Sharif. "Visually analyzing students' gaze on c++ code snippets". In: *2019 IEEE/ACM 6th International Workshop on Eye Movements in Programming (EMIP)*. IEEE. 2019, pp. 18–25 (cit. on p. 20).

[RPO+14]   M. Romanzini, E. L. Petroski, D. Ohara, A. C. Dourado, F. F. Reichert. "Calibration of ActiGraph GT3X, Actical and RT3 accelerometers in adolescents". In: *European journal of sport science* 14.1 (2014), pp. 91–99 (cit. on p. 31).

[SEBB]   M. Schweizer, T. Ertl, F. Beck, T. Blascheck. *Radial Transition Graph Comparison Tool (RTGCT)*. URL: http://www.rtgct.fbeck.com/ (cit. on p. 22).

[SED18]   B. Saket, A. Endert, Ç. Demiralp. "Task-based effectiveness of basic visualizations". In: *IEEE transactions on visualization and computer graphics* 25.7 (2018), pp. 2505–2512 (cit. on pp. 18, 23).

[Shn03]   B. Shneiderman. "The eyes have it: A task by data type taxonomy for information visualizations". In: *The craft of information visualization*. Elsevier, 2003, pp. 364–371 (cit. on pp. 17, 18).

[Sn]   I. Z. Schlueter, npm. *npm*. URL: https://www.npmjs.com/ (cit. on p. 29).

[VF]   G. Van Rossum, P. S. Foundation. *Python*. URL: https://www.python.org/ (cit. on p. 29).

[Wil]   J. Wiley. *Wiley Online Library*. URL: https://onlinelibrary.wiley.com/ (cit. on p. 24).

[WKB+21]   A. Woll, L. Klos, A. Burchartz, A. Hanssen-Doose, C. Niessner, D. Oriwol, S. C. Schmidt, K. Bös, A. Worth. "Cohort Profile Update: The Motorik-Modul (MoMo) Longitudinal Study—physical fitness and physical activity as determinants of health development in German children and adolescents". In: *International Journal of Epidemiology* 50.2 (2021), pp. 393–394 (cit. on p. 16).

[WKO+11]   A. Woll, B.-M. Kurth, E. Opper, A. Worth, K. Bös. "The 'Motorik-Modul'(MoMo): physical fitness and physical activity in German children and adolescents". In: *European journal of pediatrics* 170.9 (2011), pp. 1129–1142 (cit. on p. 39).

[WM]   J. Walke, Meta. *React*. URL: https://reactjs.org/ (cit. on p. 29).

[WWA+15]   A. Worth, A. Woll, C. Albrecht, C. Karger, N. Mewes, J. Oberger, L. Schlenker, S. Schmidt, M. Wagner, K. Boes. *MoMo-Längsschnittstudie"Physical fitness and physical activity as determinants of health development in children and adolescents": Testmanual zu den motorischen Tests und den anthropometrischen Messungen (KIT Scientific Reports; 7700)*. Vol. 7700. KIT Scientific Publishing, 2015 (cit. on p. 32).

[YW18]   C.-K. Yang, C. Wacharamanotham. "Alpscarf: Augmenting Scarf Plots for Exploring Temporal Gaze Patterns". In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI EA '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–6. ISBN: 9781450356213 (cit. on p. 22).

All links were last followed on December 03, 2021.

Declaration

I hereby declare that the work presented in this thesis is entirely
my own and that I did not use any other sources and references
than the listed ones. I have marked all direct or indirect statements
from other sources contained therein as quotations. Neither this
work nor significant parts of it were part of another examination
procedure. I have not published this work in whole or in part
before. The electronic copy is consistent with all submitted copies.

_____

place, date, signature