

Visual Analytics mit Fitnessdaten

**Studienarbeit**

des Studienganges Informatik  
an der Dualen Hochschule Baden-Württemberg Stuttgart

von

David Schönberger

10.06.2022

Matrikelnummer, Kurs

1986582, TINF19E

Ausbildungsfirma

Capgemini, Stuttgart

Betreuer

Andreas Buckenhofer

GitHub Repository

[github.com/DavidSBG/studienarbeit](https://github.com/DavidSBG/studienarbeit)

**Abstract:**

Diese Studienarbeit beschäftigt sich mit dem Thema Visual Analytics im Bezug auf von Fitness-Trackern gesammelten Daten. Dazu wird beginnend mit einer Literaturrecherche untersucht, welchen Einfluss Fitness-Tracker auf die Gesundheit der Nutzer haben können. Dabei wurde festgestellt, dass Sie, abhängig von der Nutzungsart, positive Einflüsse auf die Gesundheit haben können. Weiter wurde der Visual Analytics Prozess illustriert, welcher es ermöglicht aus einer großen Menge an komplexen Daten neue Informationen zu gewinnen. Abschließend wurden zwei Prototyp-Visualisierungen implementiert, welche Nutzern helfen können Ihre Fitness-Daten besser zu verstehen und so das Treffen besserer Entscheidungen ermöglicht.

**Abstract:**

This paper deals with the topic of visual analytics in relation to data generated by fitness trackers. Firstly, we explore how the use of fitness trackers affects one's health. The conclusion is that depending on the particular use of the tracker it can be beneficial. Further the process of visual analytics is explained. Visual analytics facilitates the generation of new knowledge based on previously existing data. Finally, two prototypes were implemented, which visualize data generated by a fitness tracker and enable the user to better understand and make decisions based on the data.

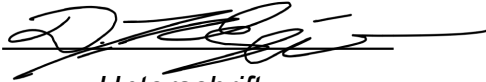
### *Erklärung*

Ich versichere hiermit, dass ich meine Studienarbeit mit dem Thema: „Visual Analytics mit Fitnessdaten“ selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Korntal-Münchingen, den 10.06.2022

*Ort*

*Datum*



*Unterschrift*

## Inhaltsverzeichnis

1.	Problemstellung .....	1
1.1.	Hintergrund .....	1
1.2.	Ziel der Arbeit .....	1
2.	Einfluss von Fitness-Trackern auf die Gesundheit.....	2
2.1.	Motivation.....	2
2.2.	Fitness Tracker in der Intervention .....	2
2.3.	Fitness Tracker zur Leistungssteigerung.....	3
2.4.	Fitness-Daten und wie sie helfen können.....	4
2.5.	Fitness-Tracking und soziale Plattformen .....	6
2.6.	Nachteile von Fitness-Trackern.....	6
3.	Visual Analytics.....	8
3.1.	Grundlagen .....	8
3.2.	Prozess .....	8
3.3.	Daten.....	10
3.4.	Visualisierungen .....	12
3.5.	Modelle.....	17
3.6.	Interaktionen .....	18
4.	FIT Protokoll .....	21
4.1.	Allgemeines.....	21
4.2.	FIT Profile.....	21
4.3.	FIT Dateiaufbau .....	24
4.4.	FIT Message Struktur.....	25
4.5.	Beispiel.....	29
5.	Visualisierungen .....	31
5.1.	Einleitung .....	31
5.2.	Die Daten .....	31
5.3.	Visualisierung 1: Trainings-Rückblick.....	33
5.4.	Visualisierung 2: Trainings-Analyse .....	42
5.5.	Weitere Visualisierungsideen .....	50
6.	Fazit.....	52
	Literaturverzeichnis .....	54

# Abbildungsverzeichnis

Abbildung 1: Verschiedene Trainingsarten.....	5
Abbildung 2: Visual Analytics Methoden.....	8
Abbildung 3: Verbreitete Visualisierungs-Möglichkeiten .....	12
Abbildung 4: Computertomographie Visualisierung .....	14
Abbildung 5: Helligkeit über Farbkategorien .....	15
Abbildung 6: Verschiedene Skalierung derselben Daten.....	16
Abbildung 7: Zusammenhang FIT Message, FIT Field und Base Type.....	21
Abbildung 8: Globales Profil und produktspezifische Profile.....	22
Abbildung 9: Globales FIT Profil CSV-Ausschnitt .....	23
Abbildung 10: Aufbau einer FIT Datei.....	24
Abbildung 11: Daten-Einträge einer FIT Datei .....	25
Abbildung 12: Zusammenspiel FIT Profil und FIT Datei .....	27
Abbildung 13: Beispiel einer gesamten FIT Datei.....	29
Abbildung 14: Visualisierung von Streckenverläufen .....	33
Abbildung 15: Farbwähler in Chromium- und Firefox-Browser .....	34
Abbildung 16: Mögliche Individualisierungen von Visualisierung 1 .....	35
Abbildung 17: Visualisierung 1 mit sieben Trainings .....	36
Abbildung 18: Beispiel von Visualisierung 1 .....	37
Abbildung 19: Auseinander gehende Farbverläufe.....	41
Abbildung 20: Punktdiagramm Visualisierung 2 .....	43
Abbildung 21: Kategorie-Diagramm Visualisierung 2 .....	43
Abbildung 22: Brushing & Linking in Visualisierung 2.....	44
Abbildung 23: Nicht gruppiertes Kategorie-Diagramm.....	45
Abbildung 24: Beispiel Visualisierung 2 .....	46
Abbildung 25: Mockup Visualisierung einzelnes Training .....	50

# Tabellenverzeichnis

Tabelle 1: Datentypen .....	10
Tabelle 2: Message Header.....	25
Tabelle 3: Definition Message Inhalt.....	26

# Abkürzungsverzeichnis

FIT	Flexible and Interoperable Data Transfer
PCA	Principle Component Analysis
SDK	Software Development Kit
CRC	zyklische Redundanzprüfung
bpm	Schläge pro Minute



# 1. Problemstellung

## 1.1. Hintergrund

Fitness Tracker sind heutzutage allgegenwärtig. 19% der Amerikaner nutzen einen Fitness Tracker und weitere 9% dokumentieren Statistiken bezüglich Ihrer Fitness mit Hilfe Ihres Smartphones[1]. Das generiert Unmengen an Fitnessdaten und stellt Nutzer und Entwickler vor eine Herausforderung. Wie lassen sich diese Daten sinnvoll nutzen? „In a world deluged by irrelevant information, clarity is power.“ [In einer Welt überschwemmt von irrelevanten Informationen ist Klarheit Macht] [2]. Das reine Ansammeln dieser Daten ist nicht ausreichend. Es braucht Klarheit in diesen Daten, um sie sinnvoll gebrauchen zu können. Visual Analytics ist ein Prozess, welcher diese Klarheit schaffen kann und es ermöglicht aus Daten neues Wissen abzuleiten.

## 1.2. Ziel der Arbeit

In dieser Arbeit soll der Einfluss von modernen Fitness Trackern auf die Gesundheit analysiert werden. Es soll einerseits überprüft werden, ob und wie sie die Gesundheit des Trägers positiv beeinflussen können. Andererseits soll ein Einblick in das Gebiet der Visual Analytics gegeben werden. Was ist Visual Analytics, wie hilft es Entscheidungen zu treffen und wie wendet man es an? Außerdem soll das Datenformat FIT (Flexible and Interoperable Data Transfer) erläutert werden, welches für die Speicherung von Fitness-Trackern erzeugten Daten verwendet wird. Schlussendlich sollen die erläuterten Informationen und gezogenen Schlüsse genutzt werden um neuartige Visualisierungen, basierend auf Daten von Fitness-Trackern, zu erstellen. Diese sollen dem Nutzer helfen die persönliche Fitness und Gesundheit zu evaluieren, vergangenen Trainings besser zu verstehen und fundierte Entscheidungen ermöglichen, um die Gesundheit zu verbessern.

## 2. Einfluss von Fitness-Trackern auf die Gesundheit

### 2.1. Motivation

Das Verwenden eines Fitness-Trackers kann von einer Reihe an Gründen motiviert sein [3]. Diese Motivation des Nutzers kann einen starken Einfluss auf die Wirksamkeit der Nutzung des Fitness-Trackers haben [4]. Verschiedene Ziele können verschieden gut mit Hilfe eines Fitness-Trackers erreicht werden. Im Rahmen dieser Arbeit werden zwei spezielle Einsatzgebiete eines Fitness-Trackers analysiert. Die Verwendung eines Fitness-Trackers im Rahmen einer Intervention und die Verwendung eines Fitness-Trackers zur Leistungssteigerung im Profisport und für die persönliche Fitness. Die Verwendung zur Leistungssteigerung ist dabei besonders bei Profisportlern anzutreffen, welche darauf angewiesen sind, zu bestimmten Zeitpunkten eine sportliche Bestleistung abrufen zu können.

### 2.2. Fitness Tracker in der Intervention

Fitness-Tracker können im Rahmen einer Intervention eingesetzt werden. Unter einer Intervention versteht man in der Pädagogik das Eingreifen in einen laufenden Prozess, üblicherweise in ein unerwünschtes Verhalten. Dieser Eingriff besteht aus einer vorher, im Rahmen einer Diagnose ermittelten, Maßnahme [5]. Im Bereich der Gesundheit könnte beispielsweise Übergewicht diagnostiziert werden. Eine daraus resultierende Maßnahme könnte das Erhöhen der täglichen Aktivität, konkret, der Schrittzahl sein. Diese Maßnahme könnte dann mit Hilfe eines Fitness-Trackers umgesetzt werden. Der Fitness-Tracker hat dabei die Aufgabe das Bewusstsein des Nutzers bezüglich der umzusetzenden Maßnahme zu erhöhen. Im Prozess einer Verhaltensänderung ist das Erhöhen des Bewusstseins ein integral Schritt, welcher hilft, das gewünschte Verhalten zu erreichen [6]. Üblicherweise wird eine Intervention von einem Arzt oder ähnlichem eingeleitet. Der Anstoß zu der Verhaltensänderung kommt bei einer Intervention also von außerhalb und ist üblicherweise nicht selbst motiviert.

Die Wirksamkeit eines Fitness-Trackers in Interventionen ist abhängig vom Kontext der Intervention. Mit dem Ziel der Steigerung der physischen Aktivität sind Interventionen, welche nur auf Fitness-Tracker setzten, sowohl bei jungen [7] als auch bei älteren [8] Zielgruppen nicht wirksam. Die Nutzung eines Fitness-Trackers zeigt hier keinen Effekt auf die Aktivität der Nutzer. Anders sieht es aus, wenn das Tracken der Fitness mit einem sozialen Aspekt verbunden wird. Eine Verbindung aus Fitness-Tracker und einer sozialen Chat-Komponente hat einen positiven Einfluss auf die Aktivität der Nutzer [9]. Effektiv ist die Intervention auch bei Nutzern, welche im Prozess der Verhaltensänderung weiter fortgeschritten sind und schon vor der Intervention ein Wille zur Verhaltensänderung zeigten [4]. Darüber hinaus ist eine Intervention auch bei chronisch-kranken Nutzern wirksam [10]. Vermutlich da bei diesen ebenso ein Wille zur Verhaltensänderung vorhanden ist.

### 2.3. Fitness Tracker zur Leistungssteigerung

Ein weiterer Use-Case für Fitness-Tracker ist die persönliche Leistungssteigerung. Im Gegensatz zu der Intervention geht es hierbei nicht um das Verbessern eines schlechten Zustandes oder das Eingreifen in ein unerwünschtes Verhalten. Sondern um das Steigern der eigenen Leistung durch die Optimierung des Trainings. Außerdem ist Ursprung hier keine ärztliche Diagnose, sondern der Nutzer entscheidet sich selbst dazu einen Fitness-Tracker zur Leistungssteigerung zu verwenden.

Ein Fitness-Tracker kann hierbei hilfreich sein, da er wichtige Metriken misst, welche Auskunft über den Trainingsverlauf geben können. Anhand dieser Daten können vergangene Trainings analysiert und Optimierungen für zukünftige Trainings vorgenommen werden. Auch während eines Trainings können Metriken überwacht werden und darauf basierend Entscheidungen getroffen werden [3]. Auch das alleinige Aufzeichnen der Trainingsdaten kann schon hilfreich sein. Das Bewusstmachen von vergangenen Leistungssteigerung kann die Motivation steigern und somit zukünftige Leistungssteigerungen ermöglichen. In Kombination mit persönlichen Zielsetzungen können Fitness-Tracker beispielsweise als Quelle von Feedback verwendet werden, was hilft die sich gesetzten Ziele tatsächlich zu erreichen [11].

Besonders im Profi- und Extremsport ist das Tracken der Fitnessdaten zum Standard geworden. Hier geben die Daten Auskunft über beispielsweise die Leistung, Genesung oder allgemeine Gesundheit des Athleten. Sie können dann verwendet werden, um die Trainings so zu planen, dass der Athlet zum Zeitpunkt des Wettkampfs auf Hochform ist [12]. Auch hilfreich sind Fitnessdaten zur Prävention von Verletzungen. Mit Fitness-Trackern kann die Belastung eines Athleten überwacht werden und bei Anzeichen von Übertraining, welches zu Verletzungen führen kann, kann vorzeitig eingegriffen werden [13, 14].

#### 2.4. Fitness-Daten und wie sie helfen können

Fitness-Tracker können verschiedenste Daten sammeln. Dabei sind Herzfrequenz, Distanz und Höhe von den meisten Trackern unterstützt. Im Folgenden wird näher darauf eingegangen, was diese Daten aussagen und wie sie beim Training helfen können.

Die Herzfrequenz gibt an, wie oft das Herz pro Minute schlägt. Die Einheit, in der die Herzfrequenz gemessen wird, ist bpm (Schläge pro Minute). Die Herzfrequenz zu tracken ist sinnvoll, da sie als Messgröße zur Einschätzung der Trainingsintensität verwendet werden kann. Das trifft besonders bei Ausdauertrainings zu. Dazu wird zuerst die maximale Herzfrequenz ermittelt. Ein ungefährender Wert lässt sich aus dem Lebensalter in Jahren  $LA$  folgendermaßen berechnen. [15]

$$HF_{max} = (220 - LA)bpm$$

Mit Hilfe der so ermittelten maximalen Herzfrequenz kann nun zu jedem Moment im Training die Intensität ermittelt werden. Dazu wird berechnet, wieviel Prozent die momentane Herzfrequenz von der maximalen Herzfrequenz entspricht. Bei einer maximalen Herzfrequenz von beispielsweise 200, entspricht eine Herzfrequenz von 160 im Training einer Intensität von 80 Prozent. Trainings können dann in verschiedene Intensitätsstufen eingeteilt werden, welche verschiedene Ziele erreichen können. [16]

Neben der Ermittlung der Intensität über die Herzfrequenz kann sie auch über die Laktatkonzentration im Blut festgestellt werden. Solche Messungen sind aber von handelsüblichen Fitness-Trackern heutzutage nicht unterstützt. [15]

	WSA- Training	GA 2- Training	KA 2- Training	GA 1- Training	KA 1- Training	KOMP- Training
Z I E L	<ul style="list-style-type: none"> <li>• Ausprägung der wettkampfspezifischen Ausdauerfähigkeit</li> </ul>	<ul style="list-style-type: none"> <li>• Entwicklung der Grundlagenausdauerfähigkeit</li> <li>• Erhöhung der aerob/ anaeroben Leistungsfähigkeit</li> </ul>	<ul style="list-style-type: none"> <li>• Entwicklung der aerob/ anaeroben Kraftausdauerfähigkeit</li> </ul>	<ul style="list-style-type: none"> <li>• Stabilisierung und Entwicklung der Grundlagenausdauerfähigkeit</li> <li>• Erhöhung der aeroben Leistungsfähigkeit</li> </ul>	<ul style="list-style-type: none"> <li>• Entwicklung und Stabilisierung der aeroben Kraftausdauerfähigkeit</li> </ul>	<ul style="list-style-type: none"> <li>• Unterstützung der Wiederherstellung</li> <li>• Erhöhung der Belastbarkeit für nachfolgendes intensives Training</li> </ul>
M E T H O D E	<ul style="list-style-type: none"> <li>• Wettkampfmethode</li> <li>• Intensive Intervallmethode</li> <li>• Wiederholungsmethode</li> </ul>	<ul style="list-style-type: none"> <li>• Extensive Intervallmethode</li> <li>• Fahrtspielmethode</li> <li>• Wechselhafte Dauer-methode</li> </ul>	<ul style="list-style-type: none"> <li>• Intensive Intervallmethode</li> <li>• Wiederholungsmethode</li> <li>• Fahrtspielmethode</li> </ul>	<ul style="list-style-type: none"> <li>• Dauer-methode</li> <li>• Fahrtspielmethode</li> </ul>	<ul style="list-style-type: none"> <li>• Dauer-methode</li> <li>• Wechselhafte Dauer-methode</li> <li>• Extensive Intervallmethode</li> </ul>	<ul style="list-style-type: none"> <li>• Kürzere Dauer-methode</li> </ul>
I N T E N S I T Ä T	<ul style="list-style-type: none"> <li>• Hoch bis sehr hoch</li> <li>• Laktat: über 6,0 mmol/l</li> <li>• Herzfrequenz (HF) &gt; 90 % der HF max</li> </ul>	<ul style="list-style-type: none"> <li>• Mittel-hoch</li> <li>• Laktat: 3,0-6,0 mmol/l</li> <li>• HF 80-90 % der HF max</li> </ul>	<ul style="list-style-type: none"> <li>• Hoch</li> <li>• Laktat: 4,0-7,0 mmol/l</li> <li>• HF 75-95 % der HF max</li> </ul>	<ul style="list-style-type: none"> <li>• Niedrig-mittel</li> <li>• Laktat: 1,5-2,5 mmol/l</li> <li>• HF 65-80 % der HF max</li> </ul>	<ul style="list-style-type: none"> <li>• Mittel</li> <li>• Laktat: 2,0-3,0 mmol/l</li> <li>• HF 75-85 % der HF max</li> </ul>	<ul style="list-style-type: none"> <li>• Sehr niedrig</li> <li>• Laktat: unter 2,0 mmol/l</li> <li>• HF 60-70 % der HF max</li> </ul>

Abbildung 1: Verschiedene Trainingsarten [16]

In Abbildung 1 sind verschiedene Trainingsarten mit verschiedenen Zielen und den dazu passenden Intensitäten dargestellt. So gibt es das Wettkampfspezifische Ausdauertraining (WSA-Training), zwei Grundlagenausdauertraining (GA-Training), zwei Kraftausdauertrainings (KA-Training) und das Kompensationstraining (KOMP-Training). Jede Art von Training hat einen anderen Zweck und ist in verschiedenen Situationen sinnvoll [16]. Trainings mit einer eher niedrigen Intensität wie das GA 1-Training oder das KA 1-Training sind beispielsweise besser dazu geeignet die aerobe Leistungsfähigkeit zu verbessern.

Trainings mit einer höheren Intensität wie das GA 2-Training oder das KA 2-Training hingegen, verbessern die aerob/anaerobe Leistungsfähigkeit. [16]

## 2.5. Fitness-Tracking und soziale Plattformen

Fitness-Tracker allein haben das Potential Nutzer zu motivieren und können für das Training hilfreiche Daten sammeln. Häufig werden Fitness-Tracker jedoch in Kombination mit sozialen Plattformen verwendet, welche den Nutzern weitere Anreize liefern sollen Ihre Ziele zu erreichen oder gar zu übertreffen. Solche Plattformen nutzen Techniken zur Verhaltensänderung. Darunter am prominentesten Selbstüberwachung, Performance-Feedbacks und Zielsetzungen, sowie Stress-Management, das Aufschauen zu Vorbildern und Simulationen von motivierenden Gesprächsführungen [17]. Das Einbinden von Fitness-Trackern in soziale Plattformen hat einerseits den Vorteil, dass die Nutzer den Fitness-Tracker zuverlässiger verwenden und weniger Nutzer aufhören den Tracker zu verwenden [18]. Andererseits konnte auch beobachtet werden, dass Nutzer im Vergleich zum alleinigen Trainieren bei der zusätzlichen Nutzung von sozialen Plattformen länger physisch aktiv sind. Das ist sowohl bei kooperativen als auch bei kompetitiven sozialen Plattformen der Fall [19].

## 2.6. Nachteile von Fitness-Trackern

Zweifellos sind Fitness-Tracker essenziell für Profi- und Extremsportler. Der Gebrauch für Hobbysportler in der Freizeit kann aber auch von Nachteil sein. Beispielsweise kann ein zu starker Fokus auf die getrackten Fitnessdaten den Spaß des Sports nehmen oder zu schlechter Laune während des Sports führen, wenn man gesetzte Ziele mal nicht erreicht und das sofort mitgeteilt bekommt [3]. Auch die Interpretation der gesammelten Daten kann zu Schwierigkeiten führen. Die isolierte Information über beispielsweise Geschwindigkeit, Herzfrequenz oder zurückgelegte Strecke ist für manche Nutzer nicht hilfreich, da Vergleichswerte fehlen oder unklar ist welche Schlüsse aus den Daten gezogen werden sollen [3].

Ein weiterer Faktor, der gegen Fitness-Tracker spricht, ist die Genauigkeit der getrackten Daten. Die Genauigkeit variiert von Hersteller zu Hersteller, ist abhängig

von dem Nutzer und ist bei jeder zu trackenden Metrik unterschiedlich. Die Genauigkeit von Entfernungsmessungen via GPS ist beispielsweise zufriedenstellend [20], wobei die Genauigkeit von optischer Pulsmessung eine signifikante Ungenauigkeit aufweist [20, 21]. Gleiches gilt für die Genauigkeit bei der Schlafüberwachung [22]. Bei der Auswahl und Nutzung eines Fitness-Trackers gilt es also sich über die Genauigkeit des speziellen Produktes zu informieren und sich bewusst zu sein, dass die getrackten Daten eine gewisse Ungenauigkeit aufweisen können.

Eine weitere Sorge, die Nutzer und potenzielle Nutzer von Fitness-Trackern haben hat mit der Sicherheit Ihrer Daten zu tun. Fitnessdaten enthalten besonders zu schützende Daten über die Gesundheit des Nutzers sowie kritische personenbezogene Daten wie den Standort. Daher besteht bei der Nutzung von Fitness-Trackern eine besondere Angst vor dem Diebstahl und der Überwachung dieser Daten [23]. Besonders zu berücksichtigen ist dieser Punkt bei der zusätzlichen Verwendung von sozialen Plattformen, da ein Teil der Daten hier an Dritte weitergegeben wird und möglicherweise frei zugänglich ist.

### 3. Visual Analytics

#### 3.1. Grundlagen

Visual Analytics beschreibt das Erstellen von abstrakten, visuellen Metaphern, welche in Interaktion mit Menschen das Entdecken neuer Informationen und das ziehen neuer Schlüsse möglich macht [24]. Visual Analytics soll helfen das Problem des Datenüberschusses zu lösen, denn Daten werden schneller generiert als sie analysiert werden können [25]. Rein automatisierte Methoden scheitern heutzutage oft an der Komplexität der zu analysierenden Daten. Es ist daher oft notwendig automatisierte Methoden und Visualisierungen mit der Intuition und dem Expertenwissen von Menschen zu kombinieren. Schlussendlich sollen so Entscheidungsfindungen auf Basis des Datenüberschusses ermöglicht werden [25].

Visual Analytics ist jedoch mehr als nur Visualisierung. Visual Analytics kombiniert Visualisierung mit statistischer Analyse, Interaktion, Wahrnehmungs -und Kognitionswissenschaft, Wissensmanagement und weiteren Aspekten [26].

#### 3.2. Prozess

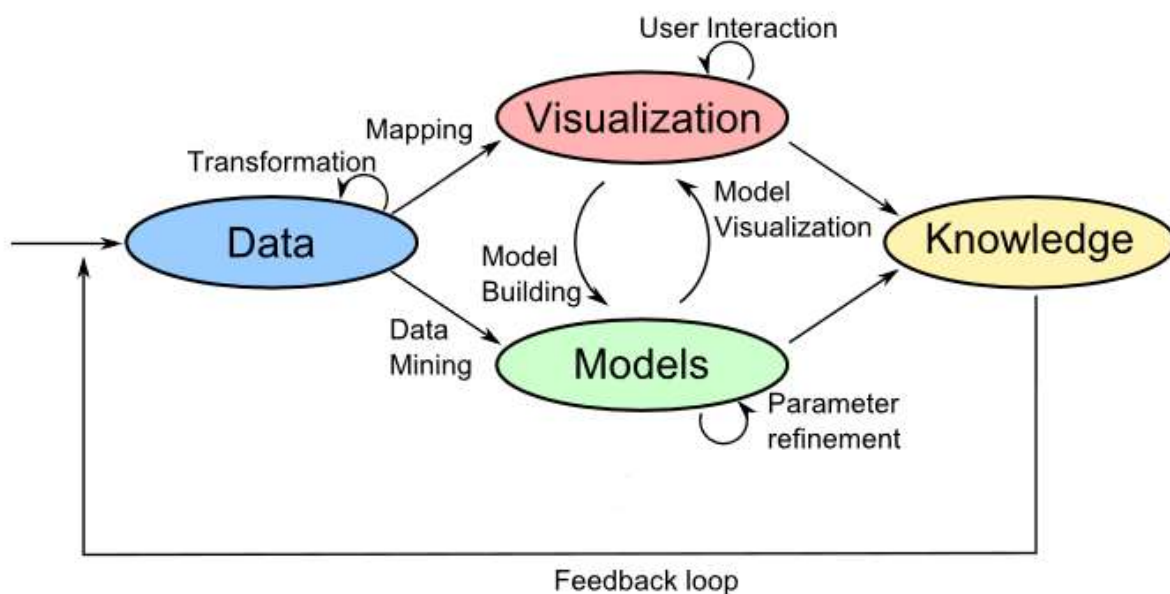


Abbildung 2: Visual Analytics Methoden [25]



Der Visual Analytics Prozess nach Keim et. al. ist in Abbildung 2 dargestellt. Er besteht aus vier zentralen Phasen. Beginnend mit der Daten-Phase. Die Daten sind Grundlage des gesamten Prozesses. Daher ist es wichtig vor Interpretation der Daten, diese bestmöglich vorzuverarbeiten. In einer der zwei anschließenden Phasen werden die vorverarbeiteten Daten nun visualisiert. Ziel der Visualisierungen ist schlussendlich das Gewinnen von neuem Wissen. Möglich ist das entweder durch eine reine Betrachtung der Visualisierung von dem Nutzer oder unterstützt von Interaktionen zwischen Visualisierung und Nutzer. Eine weitere Phase von Visual Analytics ist das Erstellen von Modellen. Das ist einerseits durch Data Mining möglich, andererseits auch durch aus den Visualisierungen gewonnenes Wissen. In Interaktion mit diesem Modell durch beispielsweise Anpassung von Parametern können wiederum weitere Einsichten gewonnen werden, welche sich durch die Visualisierung des Modells wiederum überprüfen lassen. Das insgesamt gewonnene Wissen wird schlussendlich in der letzten Phase gesammelt und innerhalb des Systems gespeichert. So neu gewonnene Daten können wiederum selbst mit dem Visual Analytics Prozess analysiert werden. [25]

### 3.3. Daten

Grundlage des gesamten Visual Analytics Prozesses sind die Daten. Daher gilt es sich vor Beginn der Analyse einen Überblick über die Daten zu verschaffen und sie gegebenenfalls vorzuverarbeiten. Vorhandene Daten können in verschiedensten Arten vorliegen. Prinzipiell kann man Daten in zwei Typen einteilen. In nominale und ordinale Daten [27]. Diese Einteilung ist in Tabelle 1 illustriert.

Nominal		Ordinal		
Aufzählungstyp	Uneingeschränkt	Binär	diskret	kontinuierlich
Farbe, Jahreszeit	Name, Rezension	0/1, ja/nein	Werte einer Zahlenmenge	1, -23, 14.056

Tabelle 1: Datentypen

Nominale Daten enthalten nicht numerische Werte. Diese können entweder Teil eines Aufzählungstyps sein, beispielsweise eine Farbe oder Jahreszeit. Hier sind die Werte Teil einer endlichen, definierten Ordnung und können deshalb sowohl nicht numerisch als auch numerisch dargestellt werden. Uneingeschränkte nominalen Daten sind ein uneingeschränkter nicht numerischer Wert, wie ein Name oder eine Rezension. Ordinale Daten hingegen enthalten rein numerische Werte. Diese können entweder binär, diskret oder kontinuierlich sein [27]. Binäre Daten sind entweder Eins oder Null, Sie haben nur zwei Möglichkeiten. Diskrete Daten sind Werte einer Zahlenmenge, beispielsweise 1 bis 12, stehend für die Monate. Kontinuierliche Daten sind nicht beschränkt, Sie können reelle Zahlen enthalten. [28]

Neben dem Typ der Daten ist es auch essenziell wichtig den Kontext der vorhandenen Daten zu verstehen. Welche Attribute beschreiben die Daten und wie wurden sie gesammelt? Nur die Kombination aus Daten und Kontext beschreibt eine Information [27].

Nachdem Typ und Kontext der Daten verstanden ist werden die Daten so vorverarbeitet, dass sie für Visualisierung und Modellbildung brauchbar sind. Hierbei ist zu beachten, dass die Vorverarbeitung kein Bias einführen und keine potenziell wertvolle Information entfernt. Außerdem sollte dem Nutzer klar kommuniziert werden, wann und wie visualisierte Daten vorverarbeitet wurden [28].

Erster Schritt der Vorverarbeitung kann die Normalisierung der Daten sein. Normalisierung wird bei kontinuierliche Daten angewandt. Dabei werden die Daten verschiedener Attribute mit unterschiedlichen Maßstäben auf einen Maßstab gebracht [29]. Beispielsweise könnte ein Datenset Daten zur Geschwindigkeit und zur Herzfrequenz eines Läufers enthalten. Angenommen Maximalwert der Geschwindigkeit ist 15km/h und Maximalwert der Herzfrequenz ist 150bpm. Hier fällt eine unterschiedliche Skalierung der Daten auf. Das kann problematisch werden, wenn aus diesen Daten ein Modell gebildet werden soll. Also auf diesen Daten ein Algorithmus wie k-nearest neighbor angewandt wird. Die unterschiedliche Skalierung führt dann zu Übergewichtung der größer skalierten Daten und verhindert somit eine sinnvolle Verwendung des Algorithmus. In solch einem Fall ist eine Normalisierung sinnvoll [29]. Ein weiterer Schritt der Vorverarbeitung kann das Erkennen von Ausreißern innerhalb der Daten sein. Es besteht immer die Möglichkeit von ungenauen Daten oder Messfehlern. Durch verschiedene Analysemethoden können diese erkannt werden. Die erkannten Ausreißer können dann entweder entfernt oder als Ausreißer markiert werden [28]. Es gibt statistische Möglichkeiten wie den Grubbs' Test zur Entdeckung von Ausreißern. Grundlage solcher Tests ist, dass die Daten annähernd normalverteilt sind [30]. Weiter gibt es komplexere Verfahren wie Clusteranalyse [31] oder Bayessche Netze [32] zur Erkennung von Ausreißern.

Abschließend in der Vorverarbeitung können die Daten noch transformiert und gemappt werden. Beispielsweise können verschiedene Attribute durch Principle Component Analysis (PCA) auf die wirklich relevanten Daten reduziert werden [28]. Auch können verschiedene Attribute zusammengeführt werden, beispielsweise Vor- und Nachname. Die Daten können dann auf ein möglicherweise schon vorhandenes Datenschema gemappt und so in Visualisierung und Modell-Bildung verwendet werden [25, 28].

### 3.4. Visualisierungen

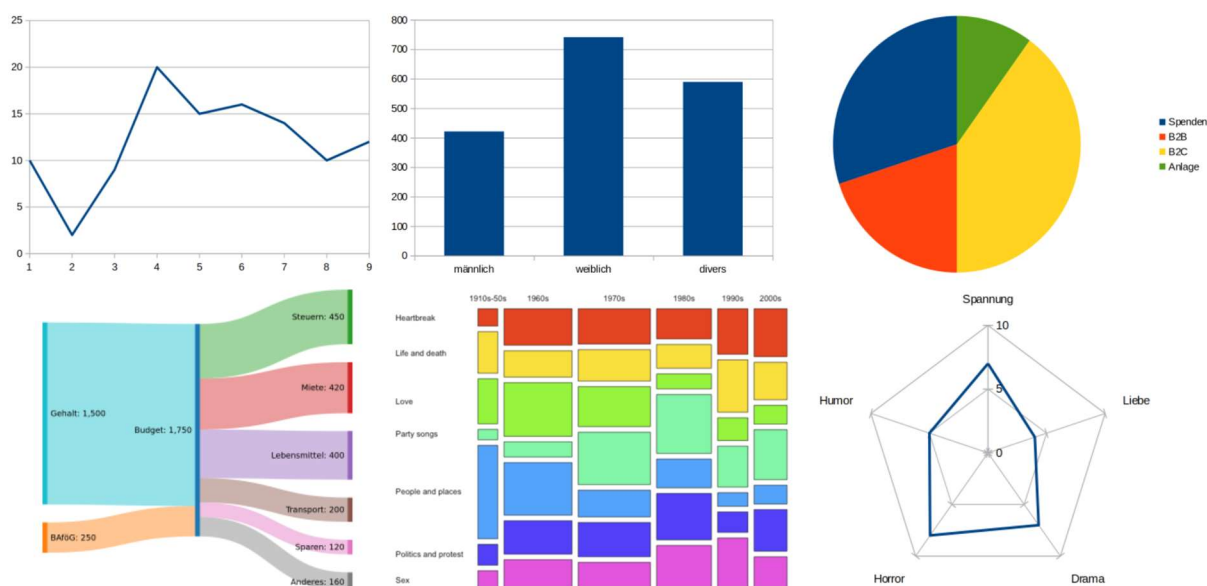


Abbildung 3: Verbreitete Visualisierungs-Möglichkeiten, von links oben nach rechts unten: (1) Liniendiagramm, (2) Säulendiagramm, (3) Kreisdiagramm, (4) Sankey-Diagramm, (5) Mosaik-Diagramm, (6) Netzdiagramm

Das Darstellen der Daten in Visualisierungen ist einer der zwei zentralen Punkte zur Wissensgewinnung in Visual Analytics. Die Visualisierung erlauben Experten in dem Fachgebiet der Daten Auffälligkeiten und Muster zu erkennen. So können sie Schlüsse ziehen und Entscheidungen treffen. Um das zu ermöglichen müssen die Visualisierung möglichst klar und verständlich sein. Sie müssen die Fähigkeit haben Tausende oder gar Millionen von Datensätzen verständlich zu repräsentieren [28].

Es gibt verschiedenste Möglichkeiten Daten zu visualisieren. In Abbildung 3 sind eine Reihe an typischen Möglichkeiten illustriert. Visualisierung (1) ist ein Liniendiagramm, welches die Daten auf einem zweidimensionalen Koordinatensystem visualisiert. Hier wird jedem X-Wert ein Y-Wert zugeordnet. Es ist auch möglich mehrere Linienverläufe in einem Diagramm darzustellen. Liniendiagramme sind besonders nützlich, um Trends über einen bestimmten Zeitraum zu visualisieren. Visualisierung (2) zeigt ein Säulendiagramm. Es visualisiert für jeden X-Wert einen Y-Wert. Im Kontrast zu dem Liniendiagramm stehen hier die X-Werte aber nicht unbedingt im Zusammenhang. X-Werte können hier beispielsweise auch nominale Beschriftungen sein. Säulendiagramme sind nützlich, um Werte verschiedener Kategorien zu visualisieren.

Beispielsweise Umsätze verschiedener Abteilungen. In Visualisierung (3) ist ein Kreisdiagramm zu sehen. Ein Kreisdiagramm visualisiert die Zusammensetzung einer Menge. Einzelne Stücke des Kreises stehen für bestimmte Untermenge und abhängig von der Größe dieser Untermenge ist die visualisierte Stückgröße. Kreisdiagramme sind nützlich um beispielsweise Zusammensetzungen von Kosten, Gewinnen oder Abstimmungen zu visualisieren. Visualisierung (4) zeigt ein Sankey-Diagramm. Ein Sankey-Diagramm ist ein Mengenflussdiagramm und visualisiert, wie eine Menge und ihre Untermengen sich aufteilen oder zusammenfließen. Dabei werden die Mengen durch Pfeile oder Balken symbolisiert, welche je nach Größe der Menge skaliert sind. Sankey-Diagramme werden verwendet um Energie-, Material- oder Geldflüsse zu visualisieren und ermöglichen es Einsparpotentiale zu erkennen. Visualisierung (5) zeigt ein Mosaikdiagramm. Ein Mosaikdiagramm ermöglicht es Datensätze mit zwei oder mehr binären Attribute zu visualisieren. Jede Kombination aus diesen Variablen wird dann ein Mosaikstück zugewiesen, dessen Größe abhängig von der Häufigkeit der Kombination im Datensatz ist. Mosaikdiagramme eignen sich besonders um multidimensionale Datensätze zu visualisieren. In Visualisierung (6) ist eine Netzdiagramm dargestellt. Ein Netzdiagramm visualisiert mehrere Werte von gleichwertigen Kategorien auf jeweils einer Achse, welche im Zentrum des Diagramms zusammenkommen. Netzdiagramme werden häufig verwendet um zwei Zahlenserien anhand festgelegter Kriterien zu vergleichen. Die Serien können entweder in zwei separaten Netzdiagrammen dargestellt werden oder zusammen in einem, wobei auf die Farbgebung geachtet werden muss.

Es ist weiter möglich verschiedene Visualisierungen zusammen zu führen und ein sogenanntes Verbunds-Diagramm zu erstellen. Es ist beispielsweise denkbar Linien- und Säulendiagramme zu verbinden.

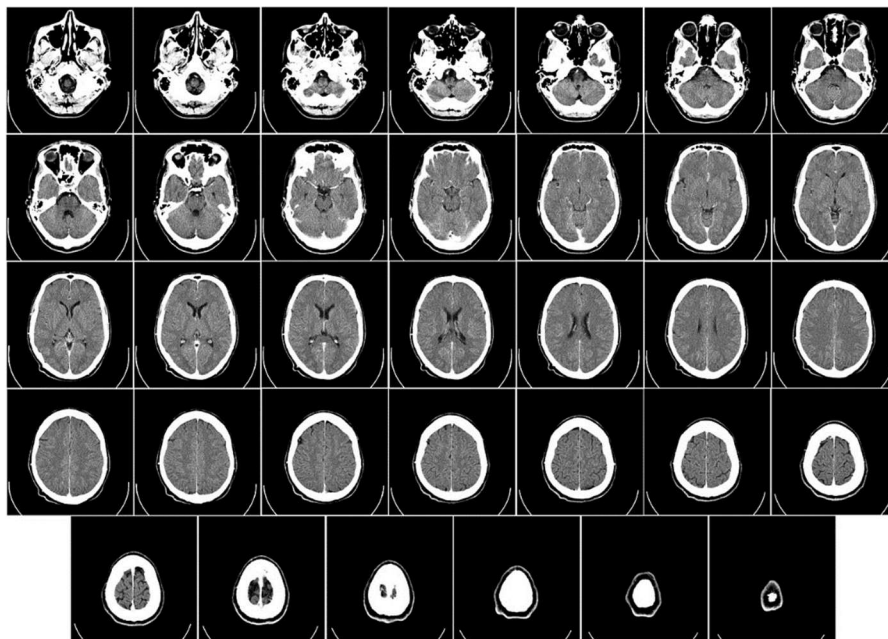


Abbildung 4: Computertomographie Visualisierung

Neben den in Abbildung 3 dargestellten Visualisierungen können auch komplexere auf die spezifischen Daten zugeschnittene Visualisierungen verwendet werden. Beispielsweise bei der Computertomographie gesammelte Daten werden in extra dazu entwickelten Visualisierungen dargestellt. In Abbildung 4 ist eine solche Visualisierung zu sehen.

Ob komplexe Computertomographie oder simples Liniendiagramm, bei Visualisierungen gilt es eine Reihe an Faktoren zu beachten, um eine erfolgreiche Interpretation der Daten zu ermöglichen. Einerseits muss die Visualisierung für die Betrachtenden ansprechend gestaltet sein und andererseits darf die Visualisierung selbst keinen Bias in die Daten einbringen.

Eine ansprechende Gestaltung der Visualisierung ist wichtig, damit ein Betrachter nicht von Gestaltungsdetails abgelenkt werden. Stattdessen soll er seine volle Aufmerksamkeit den präsentierten Informationen schenken können. Ein wichtiger Faktor in der Gestaltung einer Visualisierung ist die Wahl einer Farbpalette. Die minimale Anforderung an in einer Visualisierung genutzten Farbpalette ist, dass sie nicht ablenkend ist. Bestenfalls hilft sie dabei die Aufmerksamkeit der Betrachtenden zu halten und ist visuell ansprechend. Dazu gibt es einige Best-Practices. Es sollten

nur wenigen verschiedene Farbkategorien verwendet werden. Viele verschiedene Farben mit verschiedenen kategorischen Bedeutungen sind verwirrend, besonders wenn zusätzlich noch unterschiedliche Helligkeiten innerhalb der Kategorie verwendet werden.

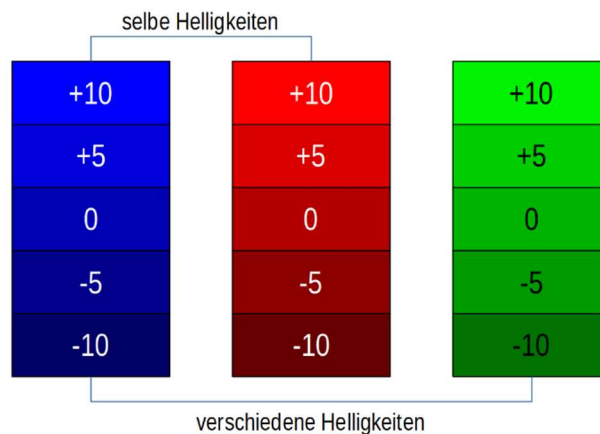


Abbildung 5: Helligkeit über Farbkategorien

Außerdem sollte die Helligkeit über verschiedene Farbkategorien hinweg einheitlich sein. Die Helligkeit einer Farbe ist oft mit dem Wert der von ihr repräsentierten Information verbunden. Wie in Abbildung 5 zu sehen sollten daher gleiche Wertunterschiede über verschiedene Kategorien gleiche Helligkeitsunterschiede haben. Schlussendlich sollten Farben gewählt werden, welche auch für Personen mit Sehschwächen zu unterscheiden sind.

Neben den rein ästhetischen Faktoren der Farbwahl ist es auch sinnvoll die Farben mit den Informationen in Zusammenhang zu bringen. Sofern möglich sollte die Farbwahl von den Informationen und Ihrem Kontext abhängen. Visualisierungen im Kontext der Forstwirtschaft könnten beispielsweise Grün oder Brauntöne verwenden. Wenn hingegen über Wasserversorgung informiert wird, wären Blautöne sinnvoll. Abschließend sollte jede Visualisierung, die über Farben Informationen vermittelt eine leicht verständliche Farblegende verwenden. [27]

Was eine gut gestaltete Visualisierung können soll stellt das Prinzip der Graphical Excellence auf [33]. Die Prinzipien besagen, eine Visualisierung soll:

- die Daten zeigen
- die Anschauenden zum Nachdenken anregen
- es vermeiden zu verzerren, was die Daten sagen
- viele Zahlen in wenig Raum zeigen
- große Datensätze in einen Zusammenhang bringen
- Vergleiche zwischen den Daten unterstützen
- die Daten in verschiedenen Detailgraden präsentieren
- einen klaren Zweck haben: Beschreibung, Exploration, Auflistung, Dekoration

Bei der Erstellung einer Visualisierung können diese Prinzipien zu Rate gezogen werden, um zu prüfen ob noch Verbesserungspotential besteht. [33]

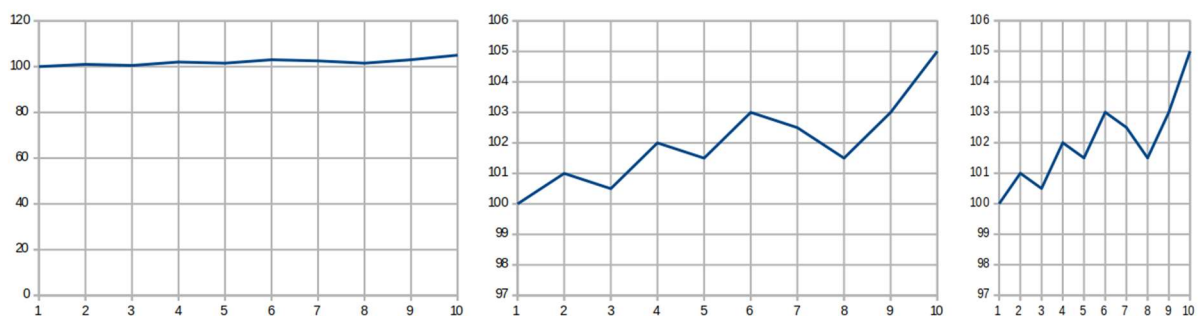


Abbildung 6: Verschiedene Skalierung derselben Daten

Neben der ansprechenden Gestaltung der Visualisierung muss auch sichergestellt werden, dass die Visualisierung keinen Bias in die Daten einbringt. Das kann beispielsweise durch eine ungünstige Wahl der Skalierung passieren. Durch Anpassen der Skalierung können große Unterschiede in den Daten klein und kleine groß erscheinen. In Abbildung 6 sind dieselben drei Daten unterschiedlich skaliert dargestellt. Um die Daten unverzerrt darstellen sollte folgendes beachtet werden. Die



Skala einer Visualisierung sollte immer bei null beginnen. So lässt sich die Veränderung eines Werts im Gesamtkontext bewerten. Außerdem sollte keine Dimension einer Visualisierung gestreckt oder gestaucht werden und es sollten keine Werte übersprungen oder zusammengefasst werden, besonders bei der Visualisierung von Daten über einen Zeitraum [28, 34].

Zusätzlich stellen das Konzept der Graphical Integrity Prinzipien auf, welche beachtet werden sollten, um die unverzerrte Visualisierung der Daten sicherzustellen. [33]

- Visualisierungen von Zahlen müssen ihren tatsächlichen Proportionen entsprechen
- Beschriftungen müssen klar und detailliert sein
- Visualisierungen dürfen keinen versteckten Zweck haben, sondern nur Schwankungen der Daten zeigen
- Es sollten bekannte Einheiten verwendet werden
- Die Zahl der visualisierten Dimensionen sollte der Zahl der Dimensionen der Daten entsprechen
- Visualisierungen sollten nichts implizieren

### 3.5. Modelle

Der zweite zentrale Punkt von Visual Analytics ist das Erstellen von Modellen. Es gibt zwei Wege, sie zu erstellen. Entweder direkt aus den vorhandenen Daten oder nach den Visualisierungen mit den dort gewonnenen Erkenntnissen. [25]

Direkt aus den Daten können Modelle durch Data Mining erstellt werden. Beispielsweise können die Daten nach Sequenzmustern durchsucht werden, dazu können Algorithmen wie der Prefix-Span Algorithmus verwendet werden [35]. Sequenzmuster sind beispielsweise nützlich bei Verkaufsdaten und können helfen das Kaufverhalten der Kunden zu verstehen. Auch Clustering-Verfahren, welche die vorhandenen Daten gruppieren können hilfreich sein und Auskunft über Zusammenhänge in den Daten geben [28].

Auch das Trainieren von überwachten Machine-Learning Modellen wie K-Nearest-Neighbor, neuronalen Netzen oder Support Vector Maschinen ist denkbar [25].

Neben dem direkten Erstellen von Modellen aus den Daten können diese auch aus Erkenntnissen der Visualisierungen hervorgehen. Beispielsweise kann erkannt werden, dass ein bestimmtes binäres Attribut immer eins ist, wenn ein anderes bestimmtes Attribut null ist. Solche Zusammenhänge können erkannt und modelliert werden.

### 3.6. Interaktionen

Die Interaktion spielt in Visual Analytics eine essenzielle Rolle. Durch die Interaktion zwischen dem Nutzer und der Visualisierung oder dem Modell können Expertise und Erfahrung des Nutzers eingebracht werden. Das ist besonders bei hochdimensionalen und komplexen Datensätzen notwendig, wo automatisierte Analysen nicht mehr ausreichen. Es kann sowohl mit Visualisierungen als auch mit Modellen interagiert werden [25].

Interaktionen mit Visualisierungen können verschiedene Aufgaben erfüllen. Diese lassen sich nach Shneiderman in die sieben Interaktions-Kategorien Überblick, Zoom, Filtern, Details auf Anfrage, In Beziehung bringen, Historie und Extrahieren unterteilen [36].

Interaktionen, die einen Überblick verschaffen sollen, erlauben es die Daten auf einem hohen, abstrakten Level zu überblicken. Dazu kann in den Daten durch Scrollen navigiert werden und es kann zwischen verschiedene Darstellungsweisen gewechselt werden. [36]

Interaktionen in der Kategorie Zoomen erlauben dem Nutzer interessante Abschnitte der Daten genauer zu betrachten. Dazu wird ein Teil der Visualisierung vergrößert. Dabei muss beachtet werden, in wieviel Dimensionen der Visualisierung gezoomt wird. Es kann sinnvoll sein das Zoomen auf eine Dimension zu beschränken. [36]

Interaktionen die Filtern sind dazu da uninteressante Daten herauszufiltern. Das wird durch das Einschränken der Attributwerte umgesetzt und kann dem Nutzer beispielsweise über Slider, Buttons oder Eingabefelder zu Verfügung gestellt werden. [36]

Details auf Anfrage beschreibt Interaktionen, die dem Nutzer zusätzlich zu den standardmäßig sichtbaren Informationen mehr Details geben. Das kann durch Auswählen von individuellen Datensätzen und anschließendem Anzeigen von den Details in einem Pop-Up realisiert werden. [36]

Interaktionen die Informationen in Beziehung bringen erlauben es Verbindungen innerhalb der Daten zu sehen und Werte dieser Daten zu vergleichen [28]. Dazu kann beispielsweise nach Auswählen eines Datensatzes die Visualisierung die angezeigten Daten auf mit der Auswahl in Verbindung stehenden beschränken. [36]

Interaktionen der Kategorie Historie erlauben es zu einem früheren Zeitpunkt durchgeführte Interaktionen zu dokumentieren, sie rückgängig zu machen oder sie zu wiederholen. So kann ein Fehler rückgängig gemacht und durch mehrfaches Interagieren das gewünschte Ergebnis erzielt werden. [36]

Das Extrahieren erlaubt es ein durch andere Interaktionen erzielttes Ergebnis zu markieren und zu exportieren. Das ist möglich durch direktes Exportieren des markierten Datensatzes in eine separate Datei oder durch Exportieren der Einstellungen und durchgeführten Interaktionen, welche zu dem gewünschten Ergebnis führen. [36]

Neben der Interaktion mit Visualisierungen kann in Visual Analytics auch mit den generierten Modellen interagiert werden. Das ist möglich durch Anpassen Ihrer Parameter [25]. Beispielsweise kann bei der Suche nach Sequenzmustern mit dem Prefix-Span Algorithmus der minimal notwendige Support angepasst werden [35]. Eine andere Möglichkeit ist das Anpassen des k-Werts bei Verwendung des K-Nearest-Neighbor Algorithmus. Neben Anpassung der Parameter kann einem vorhandenen Modell auch ein eigener Datensatz übergeben werden, um das Modell zu testen.

Allgemein lässt sich für Interaktionen noch sagen, dass direkte Interaktionen immer indirekten Interaktionen vorzuziehen sind. Das bedeutet die Interaktion soll im Rahmen der sowieso vorhanden Darstellung stattfinden. Denn es fällt leichter direkt mit der Darstellung zu interagieren statt separate Eingabefelder, Slider, Kästchen oder Unter-Menüs zu verwenden. [28]

## 4. FIT Protokoll

### 4.1. Allgemeines

Das FIT Protokoll wird verwendet um von Sport, Fitness und Gesundheitsgeräten stammende Daten zu speichern und sie zu teilen. Dazu definiert das FIT Protokoll sogenannte FIT Messages, eine Reihe an Vorlagen zur Speicherung der Daten. Diese Vorlagen können mit Daten gefüllt und in FIT-Dateien gespeichert werden. Die Struktur dieser Dateien ist ebenfalls im FIT Protokoll spezifiziert. Insgesamt besteht das FIT Protokoll aus drei Teilen. Der FIT Dateistruktur. Einer Liste aus FIT Messages, den Vorlagen, gruppiert in sogenannten FIT Profilen und dem Software Development Kit (SDK). [37]

### 4.2. FIT Profile

FIT Profile enthalten Definitionen zu den FIT Messages. Sie stellen die Vorlagen dar, welche später in der tatsächlichen Datenspeicherung verwendet werden. In einem FIT Profil werden System Konfigurationen, FIT Messages, FIT Fields und Base Types definiert.

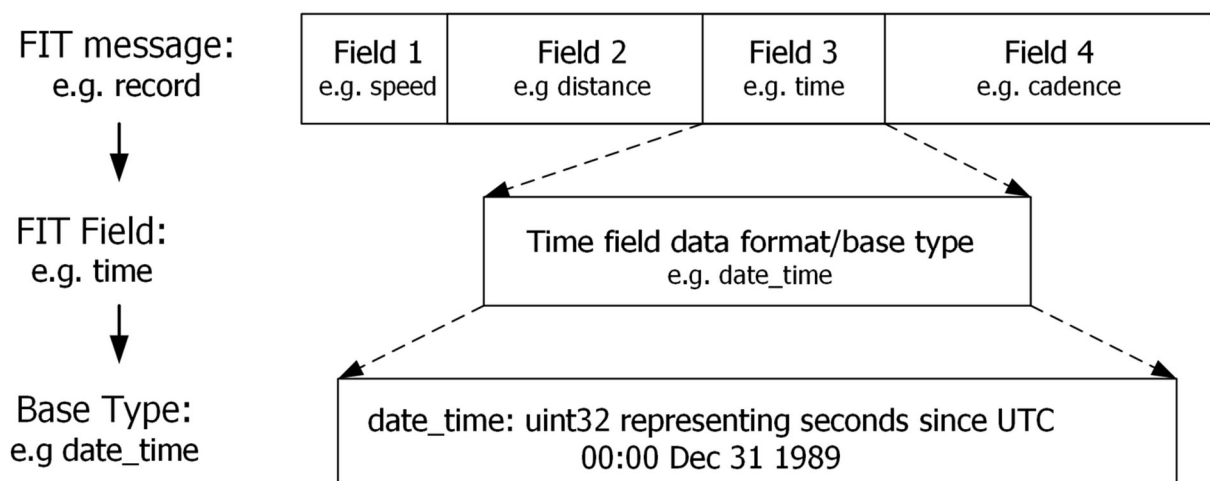


Abbildung 7: Zusammenhang FIT Message, FIT Field und Base Type [37]

Eine FIT Message steht mit einem FIT Field und einem Base Type in dem in Abbildung 7 zu sehenden Zusammenhang. Eine Fit Message besteht aus mehreren FIT Fields. Jedes FIT Field enthält wiederum eine bestimmte Information. Diese wird in einem Base Type gespeichert, welcher die zu verwendende Variable definiert. Beispielsweise enthält die FIT Message record mehrere FIT Fields. Das FIT Field time enthält beispielsweise die Informationen über die Zeit. Um diese Information zu speichern wird der Base Type date\_time verwendet, welcher schlussendlich in einem 32 Bit großen unsigned integer gespeichert wird.

Neben Informationen zu FIT Messages, FIT Fields und Base Types enthält ein FIT Profile noch System Konfigurationen. Diese beschreiben Parameter des Systems wie die Byte-Endianness.

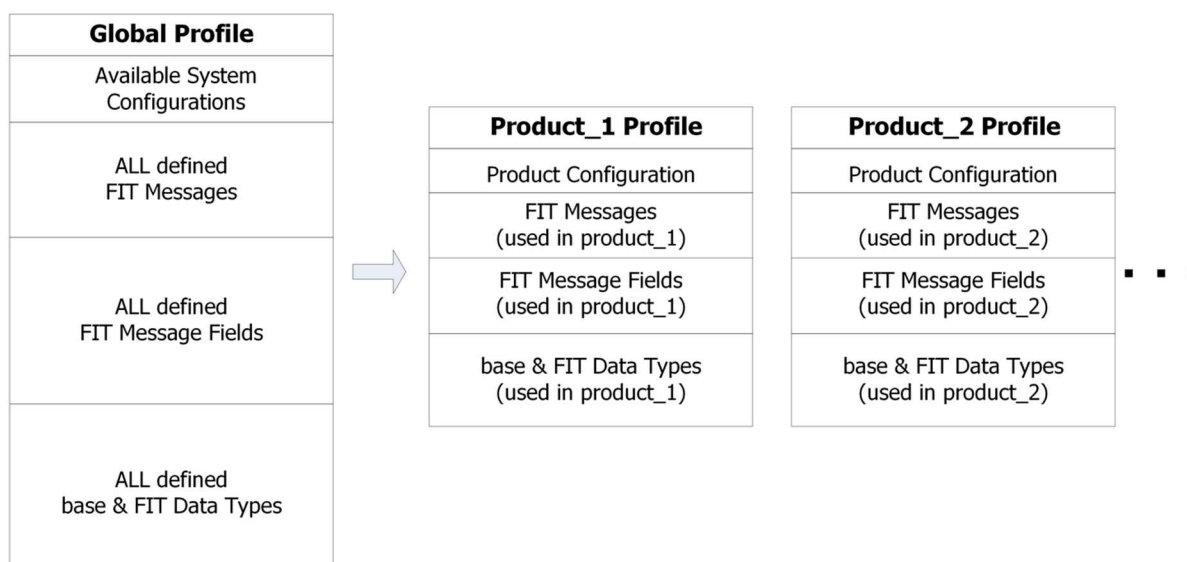


Abbildung 8: Globales Profil und produktspezifische Profile [37]

Es gibt, wie in Abbildung 8 zu sehen, zwei Arten von Profilen. Das globale Profil und produktspezifische Profile. Das globale Profil enthält Definitionen für alle von Garmin International, Inc. unterstützten System Konfigurationen, FIT Messages, FIT Fields und Base Types. Produktspezifische Profile enthalten nur eine Untermenge der in dem globalen Profil enthaltenen Definitionen. Sie beschränken sich auf die für das Produkt notwendigen Definitionen. Beispielsweise braucht ein Produkt ohne Herzfrequenzmessung keine FIT Field, welches die Herzfrequenz beschreibt.

Sowohl FIT Messages, Fit Fields als auch Base Types eines Profils sind alle nummeriert und benannt. So können andere Teile des Protokolls auf diese Definitionen in den Profilen verweisen.

Message Name	Field Def #	Field Name	Field Type	Array	Components	Scale	Offset	Units	Bits Accumulate
ACTIVITY FILE MESSAGES									
activity	253	timestamp	date_time						
	0	total_timer_time	uint32			1000	s		
	1	num_sessions	uint16						
	2	type	activity						
	3	event	event						
	4	event_type	event_type						
	5	local_timestamp	local_date_time						
	6	event_group	uint8						
session	254	message_index	message_index						
	253	timestamp	date_time				s		
	0	event	event						
	1	event_type	event_type						
	2	start_time	date_time						
	3	start_position_lat	sint32					semicircles	
	4	start_position_long	sint32					semicircles	
	5	sport	sport						
	6	sub_sport	sub_sport						
	7	total_elapsed_time	uint32			1000	s		
	8	total_timer_time	uint32			1000	s		
	9	total_distance	uint32			100	m		
	10	total_cycles	uint32				cycles		
		total_strides	uint32				strides		
		total_strokes	uint32				strokes		
	11	total_calories	uint16				kcal		
	13	total_fat_calories	uint16				kcal		
	14	avg_speed	uint16		enhanced_avg_speed	1000	m/s		16
	15	max_speed	uint16		enhanced_max_speed	1000	m/s		16

Abbildung 9: Globales FIT Profil CSV-Ausschnitt [37]

Das globale FIT Profil ist in dem FIT SDK als CSV-Datei enthalten und enthält alle Definitionen. In Abbildung 9 ist ein Ausschnitt des Profils zu sehen. Man sieht Links den Namen der FIT Message, rechts daneben alle Felder dieser Message und daneben den Base Type der Felder. Darüber hinaus enthält die Datei noch Metadaten über die Einträge wie die Skalierung oder die Einheit. Die in Abbildung 9 dargestellte Tabelle ist nur ein Ausschnitt des gesamten FIT Profils. Es gibt beispielsweise noch Teile, welche den FIT Message Namen Ihre IDs zuordnen.

Falls Hersteller Informationen speichern wollen, welche nicht durch die Definitionen in dem globalen FIT Profil unterstützt sind, haben sie die Möglichkeit eigene, produktspezifische Fit Messages, FIT Fields oder Base Types zu definieren. [37]

#### 4.3. FIT Dateiaufbau

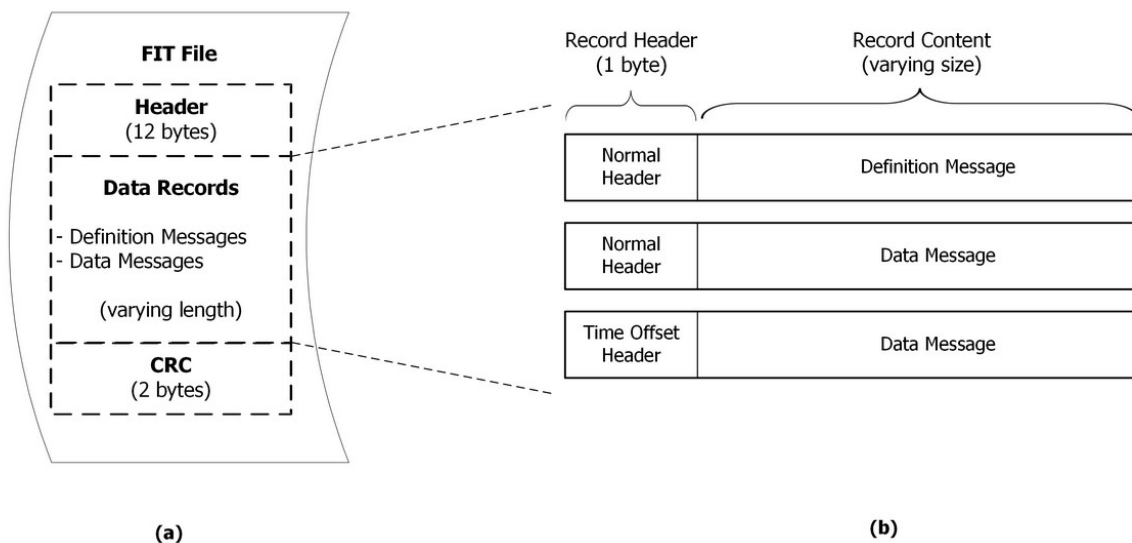


Abbildung 10: Aufbau einer FIT Datei [37]

Eine Datei im FIT Format besteht, wie in Abbildung 10 zu sehen, aus drei Teilen. Einem Header, den Daten-Einträgen und einer zyklische Redundanzprüfung (CRC). Der Header macht die ersten 12 oder 14 Bytes einer Fit Datei aus. Er enthält allgemeine Infos über die Datei. Darunter die verwendete Version des Protokolls, die Version des Profils und die Größe der Daten-Einträge. Die letzten zwei Bytes einer FIT Datei enthält die CRC, welche aus der vorgehenden FIT Datei berechnet wird. Anhand der CRC kann überprüft werden, ob die FIT Datei Fehler enthält. Zwischen Header und CRC befinden sich die Daten-Einträge, welche keine festgelegte Länge haben. Diese Einträge enthalten die tatsächlichen Informationen. [37]



#### 4.4. FIT Message Struktur

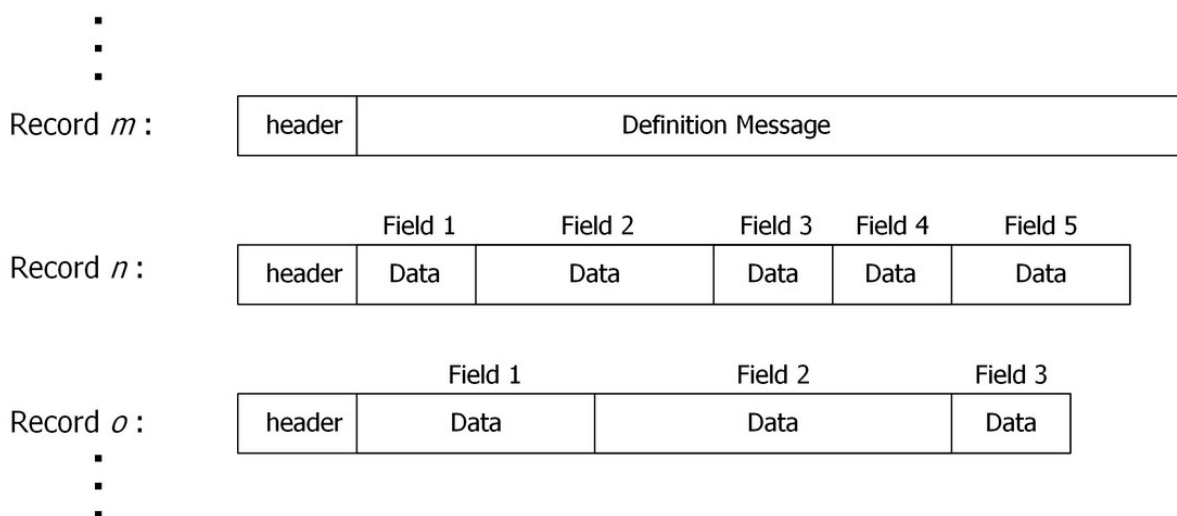


Abbildung 11: Daten-Einträge einer FIT Datei [37]

Jeder Daten-Eintrag in einer FIT Datei ist eine FIT Message. Es gibt zwei Arten von FIT Messages. Definition Messages und Data Messages. Beide bestehen, wie in Abbildung 11 zu sehen, aus einem Header und dem Inhalt. Der Header ist ein Byte groß und enthält einige allgemeine Informationen über die FIT Nachricht. Der Header ist für Definition Message und Data Message derselbe. Er ist wie in Tabelle 2 zu sehen aufgebaut. [37]

##### 4.4.1. Message Header

Bit	Wert	Beschreibung
7	0	Header Bit
6	0	Typ der Nachricht
5	0	Entwicklungs-Flag
4	0	Reserviert
0-3	0	Local Message Type

Tabelle 2: Message Header [37]

Bit 7, das Header Bit, beschreibt um welchen Typ von Header es sich handelt. Üblicherweise handelt es sich um einen normalen Header, in diesem Fall ist der Wert gleich Null und die in Tabelle 2 beschriebenen Bedeutungen gelten. Bei einem Wert von Eins handelt es sich um einen speziellen, komprimiert Timestamp-Header. Auf den hier aber nicht eingegangen wird.

Bit 6 beschreibt den Typ der auf den Header folgenden Nachricht. Bei einem Wert von Null handelt es sich um eine Definition Message, bei einem Wert von Eins um eine Data Message.

Bit 5 kann von Entwicklern gesetzt werden, hat aber sonst standardmäßig den Wert Null.

Bit 4 ist reserviert und immer Null.

Bits 0 bis 3 enthalten den Local Message Type, welcher eine Verbindung zwischen Definition Message und Data Message herstellt. Alle einer Definition Message mit dem Local Message Type 0001 zugehörigen Data Messages haben denselben Local Message Type 0001. [37]

#### 4.4.2. Definition Message Inhalt

Der Inhalt einer Nachricht unterscheidet sich zwischen Definition und Data Messages. Der Inhalt von Definition Messages ist wie in Tabelle 3 zu sehen aufgebaut.

Byte	Länge	Beschreibung
0	1 Byte	Reserviert, immer 0
1	1 Byte	Architektur
2-3	2 Bytes	Global Message Number
4	1 Byte	Anzahl der Felder
5 – (4 + Felder * 3)	3 Byte pro Feld	Felder-Definitionen

Tabelle 3: Definition Message Inhalt [37]

Das Architektur Byte 1 beschreibt, ob die System-Architektur Big oder Little Endian verwendet. Null steht für Little Endian, Eins für Big Endian. Alle Daten in der Definition Message und den folgenden Data Messages folgen diesem Format.

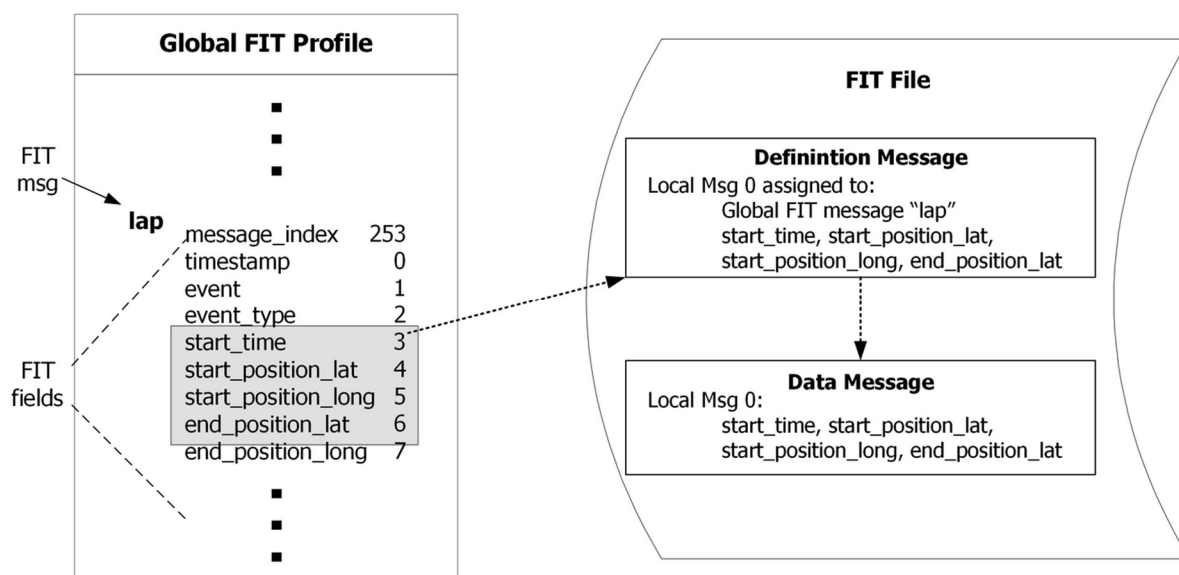


Abbildung 12: Zusammenspiel FIT Profil und FIT Datei [37]

Die Global Message Number verknüpft die Definition Message und die mit ihr verbundenen Data Messages mit einer, in einem Profil definierten globalen FIT Message.

Über die Anzahl der Felder wird definiert, wie viele Felder der verknüpften globalen FIT Message verwendet werden. Es müssen nicht alle im Profil für die FIT Message definierten Felder verwendet werden, es kann auch nur eine Untermenge genutzt werden.

Die Felder-Definitionen beschreiben konkret, welche Felder der verknüpften globalen Fit Message verwendet werden sollen. Dazu wird die im Profil definierte Nummer des Felds, die Größe und ihr Base Type angegeben.

In Abbildung 12 ist abstrakt das Zusammenspiel zwischen FIT Profil, Definition Message und Data Message in einer FIT Datei zu sehen. Die Definition Message enthält den Local Message Type, hier 0, welcher die Verknüpfung zwischen Definition und Data Message erlaubt.

Weiter enthält die Definition Message die Global Message Number, hier mit Ihrer Beschreibung lap, welche die Verknüpfung zu dem zugehörigen Eintrag im FIT Profil ermöglicht. Aus den im FIT Profil definierten Feldern wählt nun die Definition Message eine Untermenge, hier start\_time, start\_position\_lat, start\_position\_long und end\_position\_lat. Diese Felder werden nun in allen Verknüpften Data Messages verwendet. [37]

#### 4.4.3. Data Message Inhalt

Nachdem ein Definition Message die verwendeten Felder definiert hat, können Data Messages geschrieben werden. Diese sind durch den Local Message Type miteinander verbunden. Nach dem Header einer Data Message folgt sofort Ihr Inhalt. Aufgeteilt ist dieser in die definierten Felder. Alle haben eine gewisse Länge und einen definierten Datentyp. [37]

## 4.5. Beispiel

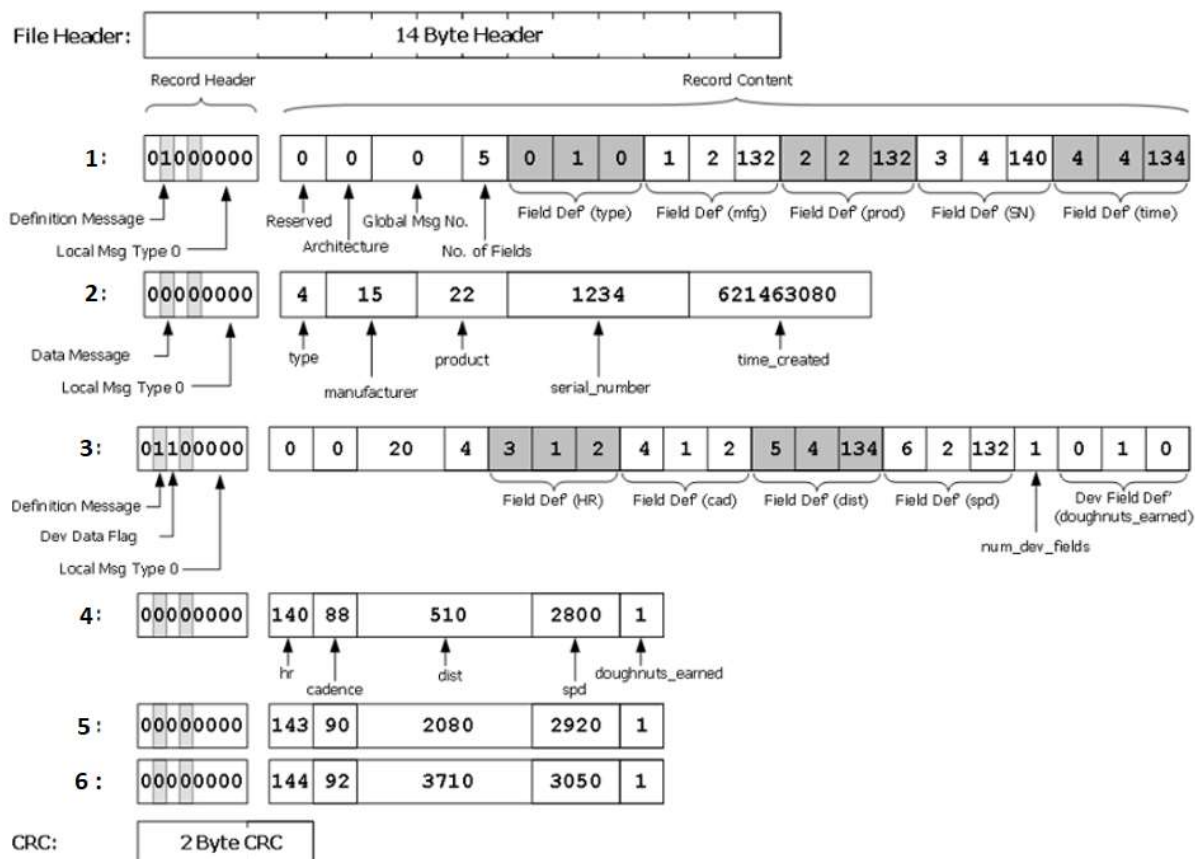


Abbildung 13: Beispiel einer gesamten FIT Datei [37]

In Abbildung 13 ist ein Beispiel einer gesamten FIT Datei zu sehen. Sie beginnt wie beschrieben mit dem Datei Header und endet mit der CRC. Dazwischen sind die Einträge zu sehen. Bei Eintrag 1 und Eintrag 3 handelt es sich, wie in dem markierten Bit zu sehen, um Definition Messages. Eintrag 1 ist eine Definition Message mit dem Local Message Type 0000. Sie verweist mit ihrer Global Message Number 0 auf FIT Message 0 im zugehörigen Fit Profil. Dann wird spezifiziert, wie viele Felder der FIT Message verwendet werden sollen. Anschließend wird jedes dieser Felder definiert. Beispielsweise das Feld *type*, welches die Nummer 0, Länge 1 und Base Type 0 hat. Base Type 0 ist hierbei ein 1 Byte langes Enum. Eintrag 2 ist eine Data Message, erkennbar an dem 0 Bit in Ihrem Header. Diese Data Message gehört zu der Definition Message aus Eintrag 1, da sie denselben Local Message Type besitzen. Nach dem

Header enthält diese Message die definierten Felder. Man erkennt, dass die Länge der Felder der in Eintrag 1 definierten Längen entspricht.

Eintrag 3 ist eine weitere Definition Message. Es fällt auf, dass sie denselben Local Message Type wie die Definition Message aus Eintrag 1 besitzt. Das ist möglich, da Local Message Types überschrieben werden können. Sobald eine Definition Message einen schon definierten Local Message Type nutzt, wird dieser überschrieben und alle kommenden Data Messages mit diesem Local Message Type verwenden die neuen Definitionen. Eintrag 4, 5 und 6 sind solche Data Messages. Sie enthalten, wie in Eintrag 3 definiert die Daten.

Hier fällt außerdem auf, dass das Entwickler-Flag gesetzt ist. Das erlaubt es Entwicklern eigene Felder hinzuzufügen. In Eintrag 3 wird das zusätzliche Feld *doughnuts\_earned* definiert, welches in Einträgen 4, 5 und 6 verwendet wird. [37]

## 5. Visualisierungen

### 5.1. Einleitung

Im Rahmen dieser Arbeit wurden zwei Prototypen zur Visualisierung von Fitnessdaten im FIT-Format implementiert. Diese Visualisierungen sollen dem Nutzer helfen Ihre Daten und somit Ihre Fitness und Gesundheit besser zu verstehen. Die Visualisierungen sind in Python entwickelt und werden in einem Jupyter Notebook bereitgestellt.

### 5.2. Die Daten

Die zu visualisierenden Daten wurden während Ausdauertrainings aufgezeichnet und enthalten folgende Informationen.

- Höhe

Die Höhe über dem Meeresspiegel in Meter.

- Schrittfrequenz

Die Schrittfrequenz wird in Umdrehungen pro Minute oder rpm gemessen und gibt an, wie oft pro Minute ein Schritt gemacht wurde.

- Distanz

Die Distanz misst die Länge der während des Trainings zurückgelegten Strecke in Metern. Sie beginnt immer bei null und steigt pro zurückgelegtem Meter.

- Herzfrequenz

Die Herzfrequenz des Nutzers wird in Schlägen pro Minute (bpm) gemessen.

- Position

Die Position des Nutzers wird über das Training hinweg getrackt und in Form von Längen- und Breitengraden gespeichert.

- Geschwindigkeit

Die Geschwindigkeit wird in Metern pro Sekunde gespeichert.

- Temperatur

Die Temperatur wird in Grad Celsius gespeichert

- Zeitstempel

Zu jedem Datenpunkt wird ein Zeitstempel bestehend aus Datum und Uhrzeit in Form YYYY-MM-DD hh:mm:ss gespeichert.

Für die implementierten Visualisierungen relevant sind die Distanz, die Herzfrequenz, die Position, die Geschwindigkeit und der Zeitstempel. Alle Fitness-Tracker, welche diese Daten unterstützen und im FIT-Format speichern, können die implementierten Visualisierungen verwenden.



## 5.3. Visualisierung 1: Trainings-Rückblick

### 5.3.1. Hintergrund

Das Ziel dieser Visualisierung ist es einen schnell verständlichen Überblick über vergangene Trainings zu ermöglichen. Die Trainings sollen auf den ersten Blick zu erkennen sein, oberflächliche Infos über die Trainings kommunizieren und somit eine Einschätzung des Trainings-Fortschritts ermöglichen. Dazu werden Informationen über Position, Distanz und Datum des Trainings visualisiert. Die Positionsinformationen werden genutzt, um, wie in Abbildung 14 zu sehen, für jedes Training ein kleines Bild des Streckenverlaufs zu visualisieren.

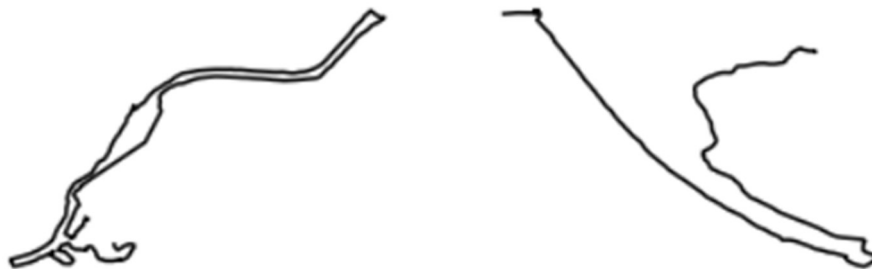


Abbildung 14: Visualisierung von Streckenverläufen

So kann anhand des Streckenverlaufs ein Training schnell zugeordnet werden. Beispielsweise können tägliche Läufe zur Arbeit anhand des sich oft wiederholenden, bekannten Verlaufs erkannt werden. Aber auch große Ereignisse wie ein Marathonlauf können an Ihrer markanten Streckenverläufen identifiziert werden. Zusätzlich zur reinen Visualisierung des Streckenverlaufs wird dieser je nach zurückgelegter Distanz eingefärbt. Dazu wird eine Farblegende bereitgestellt, welche es erlauben soll, die zurückgelegte Distanz jedes Trainings auf einen Blick direkt einzuschätzen. Außerdem werden mit der Information des Datums die Nutzer informiert, wann das Training stattgefunden hat. Ein möglicher Use-Case dieser Visualisierung ist ein Monats- oder Jahresrückblick. Ein Nutzer lädt die über Wochen, Monate oder Jahre gesammelten Daten, visualisiert diese und erkennt direkt Trends und Muster. So kann beispielsweise erkannt werden, dass die zurückgelegte Distanz über das Jahr hinweg immer weiter gestiegen ist. Oder, dass nach Erreichen eines Ziels, beispielsweise nach Laufen eines Marathons, weniger trainiert wurde.

### 5.3.2. Funktionsweise

Vor Darstellung der Visualisierung müssen die notwendigen Daten geladen werden. Da die Visualisierung mehrere Trainings visualisiert wird ein Unterordner angelegt, welcher alle zu nutzenden FIT-Dateien enthält. Aus jeder geladenen FIT-Datei werden die Positionsdaten, die Distanz und das Datum des Trainings ausgelesen. Diese müssen in allen FIT-Dateien enthalten sein. Für jedes Training wird dann der Streckenverlauf visualisiert und abhängig von der Distanz eingefärbt. Dabei werden die Trainings nach Datum sortiert, visualisiert und beschriftet. Das älteste Training befindet sich oben links, das neueste unten rechts.

Die Visualisierung erlaubt eine Reihe an Individualisierungen. Sowohl die Hintergrundfarbe, Beschriftungsfarbe als auch der Farbverlauf können über Farbwähler angepasst werden. Implementiert ist die Farbauswahl mit Jupyter Widgets, welche je nach Browser unterschiedlich aussieht.



Abbildung 15: Farbwähler in Chromium- und Firefox-Browser

In Abbildung 15 ist links die Farbauswahl in auf Chromium basierenden Browsern zu sehen und rechts die Farbauswahl in Firefox. Für von Jupyter-Widgets nicht unterstützte, beispielsweise ältere Browser ist es außerdem möglich die Farben manuell über Hex-Werte zu definieren.

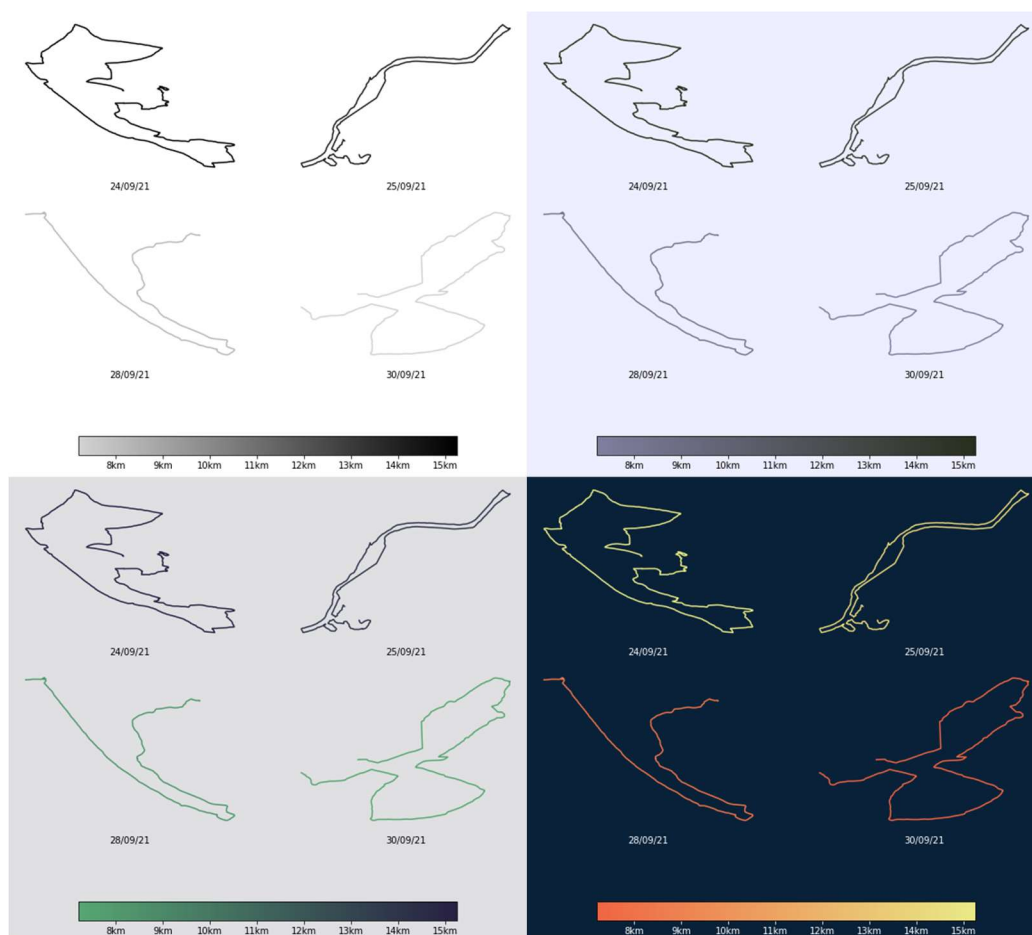


Abbildung 16: Mögliche Individualisierungen von Visualisierung 1

In Abbildung 16 sind eine Auswahl an möglichen Individualisierungen zu sehen. Bei der Individualisierung der Visualisierung ist es empfehlenswert die Hintergrundfarbe so zu wählen, dass sie einen möglichst großen Kontrast zu den Farbverlaufs- und der Beschriftungsfarbe hat. Sonst sind Streckenverläufe oder die Datumsbeschriftung nur schwer zu lesen. Bei der Wahl der Farben für den Farbverlauf sollte beachtet werden, dass die Anfangs- und End-Farbe möglichst unterschiedlich ist. Auf einen Blick sollen zwei Strecken direkt verglichen werden können. Bei einem Farbverlauf mit zwei sehr ähnlichen Farben ist das nicht möglich. Weiter ist empfehlenswert, für die End-Farbe des Farbverlaufes eine Farbe mit einem größeren Kontrast zum Hintergrund als der Kontrast zwischen Anfangs-Farbe und Hintergrund zu wählen. So stechen die langen Strecken besonders hervor.

### 5.3.3. Herausforderungen

Eine Herausforderung bei der Implementierung dieser Visualisierung war die Anordnung der Trainings. Es kann keine festgelegte Anordnung geben, da die Anzahl der zu visualisierenden Trainings variabel ist. Die Trainings werden in einem Rechteck angeordnet. Anhand der Anzahl an Trainings muss vor der Visualisierung die Länge und Breite der Visualisierung ermittelt werden. Wie viele Trainings sind in einer Reihe und wie viele in einer Spalte?

Der implementierte Ansatz versucht die Trainings immer quadratisch anzuordnen. Die Länge und Breite der Visualisierung entspricht der aufgerundeten Wurzel der Anzahl an Trainings. Besonders gut sieht die Visualisierung daher aus, wenn die Anzahl an zu visualisierenden Trainings eine Quadratzahl ist. Sonst enthält die Visualisierung, wie in Abbildung 17 zu sehen, eine freie Fläche.

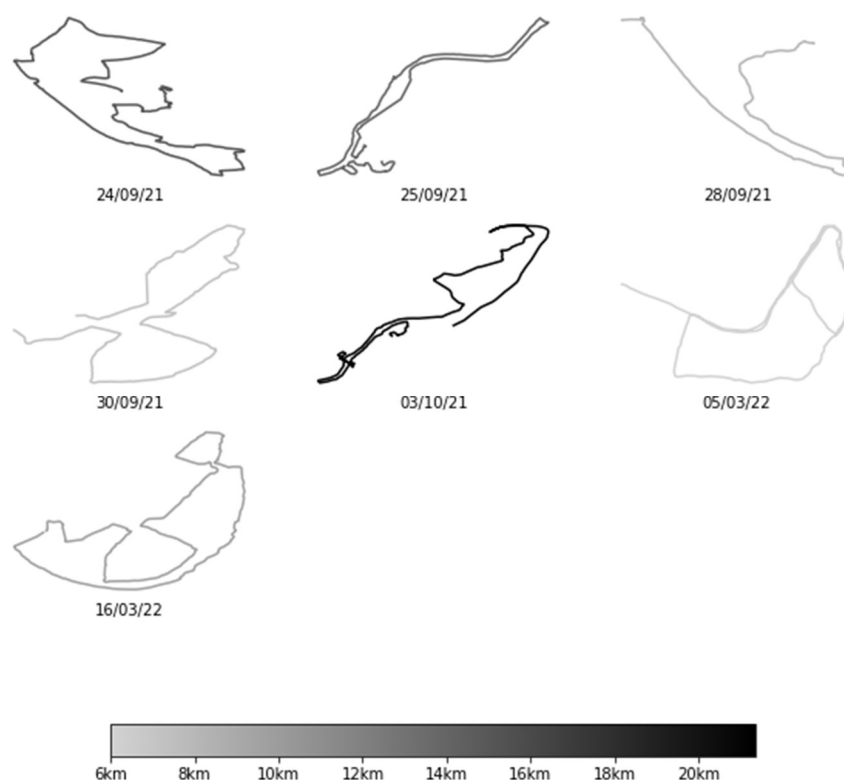


Abbildung 17: Visualisierung 1 mit sieben Trainings

Ein alternativer Ansatz wäre gewesen, anhand der Trainingsanzahl verschiedene mögliche Anordnungen zu generieren und den Nutzer eine Anordnung wählen zu lassen. Es könnte auch eine automatische Wahl getroffen werden, welche die Anordnung mit der geringsten freien Fläche wählt. Im Rahmen dieses Prototyps wurde jedoch die Anordnung in einem Quadrat gewählt, da sie schnell ein zufriedenstellendes Ergebnis liefert.

#### 5.3.4. Bewertung

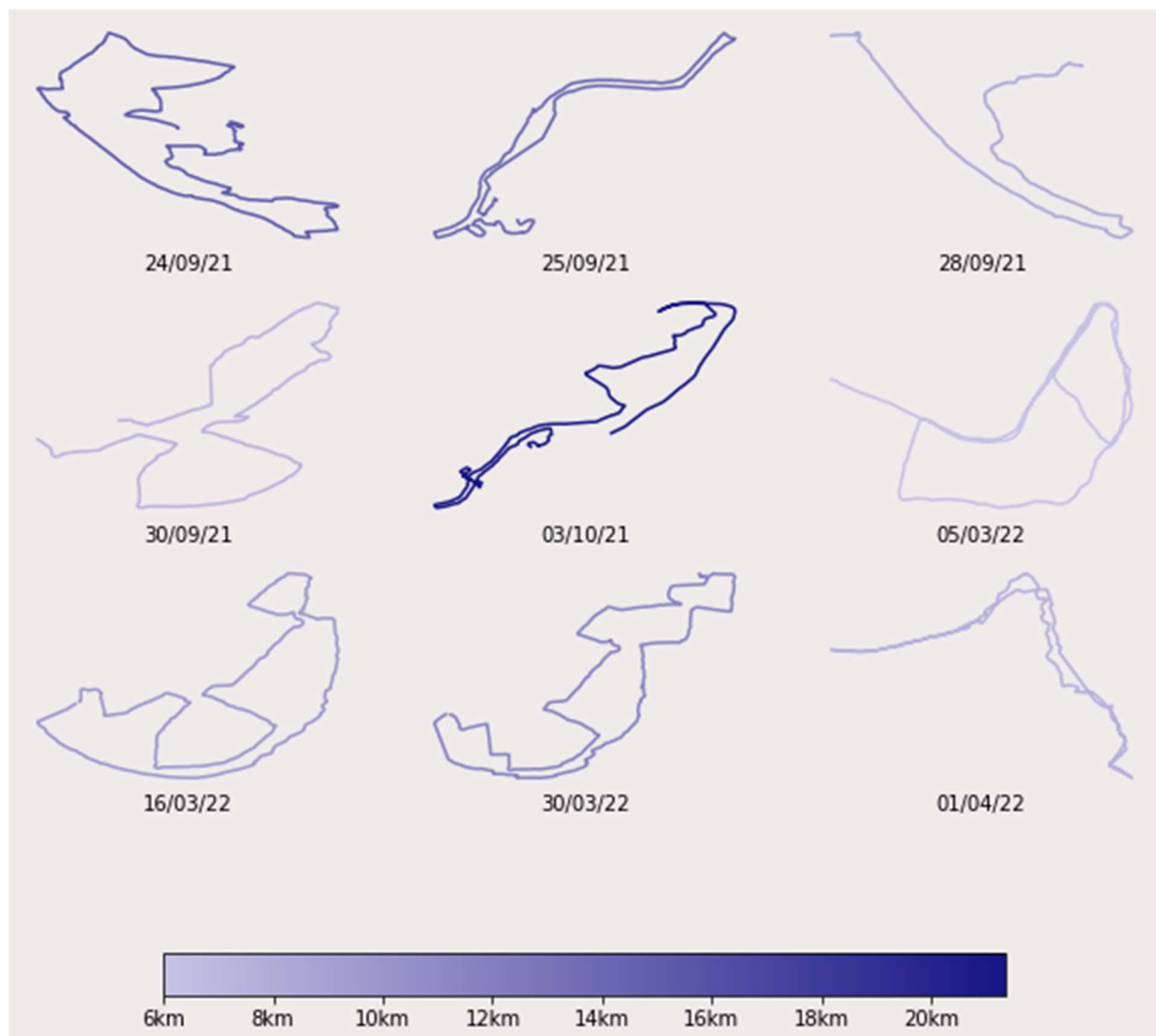


Abbildung 18: Beispiel von Visualisierung 1

In Abbildung 18 ist ein Beispiel einer individualisierten Visualisierung mit neun Trainings zu sehen. Anhand Tufte's Prinzipien der Graphical Excellence und Graphical Integrity soll die Visualisierung nun bewertet werden.

Erfüllt die Visualisierung Tufte's Prinzipien der Graphical Excellence [33]?

- Zeigt Sie die Daten?

Ja, Sie zeigt die Daten.

- Regt Sie den Betrachter zum Nachdenken an?

Ja, Sie regt den Betrachter zum Nachdenken an. Der Betrachter versucht hervorgehobene Trainings mit einer hohen Distanz zuzuordnen und kann über erkennbare Auf-/Ab-/Seitwärtstrends nachdenken.

- Vermeidet Sie es zu verzerren, was die Daten sagen?

Nein, die Daten des Streckenverlaufs werden verzerrt. Das Bild jeder Strecke hat dieselbe Größe. Kleine Strecken wirken daher ähnlich wie große Strecken. Aber, um dem entgegenzuwirken sind die Strecken nach Distanz eingefärbt. Bei einer Einfärbung nach einer anderen Metrik müsste die Skalierung des Streckenverlaufs beachtet werden.

- Zeigt Sie viele Zahlen in wenig Raum?

Ja, Sie zeigt eine Reihe an Trainings mit vielen Zahlen des Streckenverlauf und Informationen über Datum und Distanz auf einem kleinen Raum.

- Bringt Sie große Datensätze in einen Zusammenhang?

Ja, Sie bringt die Datensätze der einzelnen Trainings in einen Zusammenhang.

- Unterstützt Sie Vergleiche zwischen den Daten?

Ja, Sie erlaubt es einzelne Trainings anhand der Distanz und des Streckenverlaufes miteinander zu vergleichen.

- Präsentiert Sie die Daten in verschiedenen Detailgraden?

Nein, Sie präsentiert die Daten nur auf einem Detailgrad. Hier könnten durch Interaktionen oder durch eine Erweiterung der Visualisierung mehr Detailgrade dargestellt werden. Beispielsweise könnte man die Daten hierarchisch in Tage/Wochen/Monate aufteilen und für jede Hierarchie eine Visualisierung anbieten.

- Hat Sie einen klaren Zweck?

Ja, Sie soll vergangene Trainings darstellen und einen Vergleich dieser ermöglichen.

Erfüllt die Visualisierung Tufte's Prinzipien der Graphical Integrity [33]?

- Entsprechen Visualisierungen von Zahlen Ihren tatsächlichen Proportionen?

Nein, die Proportionen des Streckenverlaufs sind verzerrt. Aber, um dem entgegenzuwirken sind die Strecken nach Distanz eingefärbt.

- Sind Beschriftungen klar und detailliert?

Ja, die Beschriftungen sind klar und detailliert.

- Hat die Visualisierung keinen versteckten Zweck, sondern zeigt nur Schwankungen der Daten?

Ja, es werden nur die Schwankungen der Daten gezeigt.

- Werden bekannte Einheiten verwendet?

Ja, es wird nur die bekannte Einheit Kilometer verwendet.

- Entspricht die Zahl der visualisierten Dimensionen der Zahl der Dimensionen der Daten?

Ja. Die Daten sind 4-dimensional, bestehend aus einer Menge Längengraden, Breitengraden, der Distanz und des Datum des Trainings. Die visualisierten Dimensionen entsprechen den 4-Dimensionen. Der Streckenverlauf stellt Längen- und Breitengrade zweidimensional dar. Dazu kommen die Beschriftung und die Einfärbung als dritte und vierte Dimension.

- Impliziert die Visualisierungen nichts?

Nein. Die Daten implizieren, dass es sich um Fitness-Daten handelt. Im Rahmen der Studienarbeit wird dies aber explizit erwähnt. Ein entsprechender Titel oder eine Beschreibung wäre jedoch zusätzlich angebracht.

#### 5.3.5. Mögliche Weiterentwicklung des Prototyps

Bei einer Weiterentwicklung des Prototypen könnte eine Individualisierbarkeit der Anordnung ermöglicht werden. Der Nutzer könnte wählen welche Anordnung die Visualisierung haben soll. Es könnte beispielsweise gewählt werden, dass die Trainings in drei Reihen und fünf Zeilen angeordnet werden sollen. Auch komplexere Anordnungen, wie die Anordnung in einem Kreis, wären denkbar.

Dem Nutzer könnte außerdem die Möglichkeit gegeben werden die Metrik, nach welcher der Streckenverlauf eingefärbt wird, anzupassen. Statt nach der Distanz des Trainings einzufärben könnte die durchschnittliche Geschwindigkeit, Herzfrequenz oder die Höhe des Trainings verwendet werden.

Außerdem könnten bei einer Weiterentwicklung des Prototyps komplexere Farbverläufe realisiert werden. In der momentanen Implementierung sind nur sequenzielle Farbverläufe bestehend aus zwei Farben möglich. Darüber hinaus könnte man Farbverläufe bestehend aus drei oder mehr Farben erlauben oder auseinander gehende Farbverläufe integrieren.



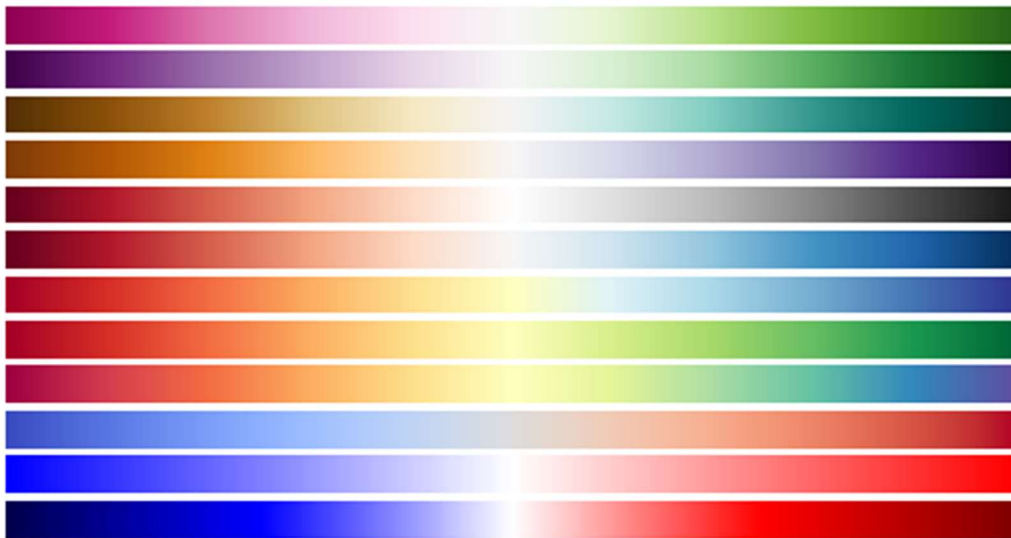


Abbildung 19: Auseinander gehende Farbverläufe [38]

Auseinander gehende Farbverläufe steigen, wie in Abbildung 19 zu sehen, monoton in Ihrer Helligkeit bis zu dem Mittelpunkt des Farbverlaufes. Ab dem Mittelpunkt sinkt die Helligkeit monoton. In auseinander gehenden Farbverläufen können nur eine oder mehrere verschiedene Farben verwendet werden. Ein solcher Farbverlauf hebt beide Extreme, sowohl die niedrigen als auch die hohen, hervor und stellt die durchschnittlichen Mittelwerte in den Hintergrund. [38]

Die im Rahmen der Bewertung nach Tufte's Graphical Excellence und Graphical Integrity festgestellten Mängel könnten behoben werden. Die Skalierung des Streckenverlaufs könnte angepasst werden, sodass Sie den tatsächlichen Proportionen entspricht. Außerdem könnte die Visualisierung erweitert werden, um mehrere Detailgrade zu zeigen und Sie könnte ausgiebiger beschrieben werden.

## 5.4. Visualisierung 2: Trainings-Analyse

### 5.4.1. Hintergrund

Diese Visualisierung soll eine tiefgehende Analyse von vergangenen Trainings erlauben. Dazu werden Informationen über die zurückgelegte Distanz, die durchschnittliche Geschwindigkeit, die durchschnittliche Herzfrequenz und den Zeitpunkt des Trainings gesammelt. Der Nutzer soll über diese Daten einen Überblick erhalten und ihm soll durch verschiedene Interaktionen erlaubt werden die Daten besser zu verstehen.

Ein Nutzer dieser Visualisierung könnte beispielsweise rückblickend überprüfen, ob sich über die vergangenen Wochen oder Monate seine durchschnittliche Geschwindigkeit in einer bestimmten Distanz verbessert oder verschlechtert hat.

### 5.4.2. Funktionsweise

Vor Darstellung der Visualisierung werden die notwendigen Daten geladen. Da die Visualisierung mehrere Trainings visualisiert, wird ein Unterordner angelegt, welcher alle zu nutzenden FIT-Dateien enthält. Aus jeder geladenen FIT-Datei werden die Distanz, das Datum, die durchschnittliche Herzfrequenz und die durchschnittliche Geschwindigkeit des Trainings ermittelt. Dazu müssen in allen FIT-Dateien Distanz, Datum, Herzfrequenz und Geschwindigkeit enthalten sein. Die durchschnittliche Herzfrequenz und die durchschnittliche Geschwindigkeit sind nicht fester Teil des FIT-Protokolls. Stattdessen werden während jedes Trainings die momentane Geschwindigkeit und Herzfrequenz getrackt. Zur Ermittlung der Durchschnittswerte werden dann ab Beginn des Trainings alle Einzelwerte addiert und schlussendlich durch die gesamte Trainingszeit geteilt. So werden Durchschnitts-Geschwindigkeit und Durchschnitts-Herzfrequenz ermittelt.

Vor der Visualisierung werden die durchschnittlichen Herzfrequenzen und Geschwindigkeiten gruppiert. Der Nutzer kann die Anzahl der Gruppen individualisieren. Standardmäßig werden vier Gruppen erstellt. Die Aufteilung der Daten in die Gruppen funktioniert folgendermaßen. Zuerst wird in jeder Kategorie der Minimalwert von dem Maximalwert abgezogen. So erhält man die Spanne der Werte in der Kategorie.

Diese Spanne wird nun in die Anzahl der festgelegten Gruppen geteilt. Jeder Wert fällt jetzt in die Spanne einer Gruppe und wird so zugeordnet. Statt den tatsächlichen Werten, beispielsweise einer durchschnittlichen Geschwindigkeit von 9.03km/h, enthält der Wert jetzt eine Gruppe, beispielsweise 8km/h-10km/h. Die Gruppierung der Distanz ist durch ein Runden auf den Nächsten vollen Kilometer realisiert.

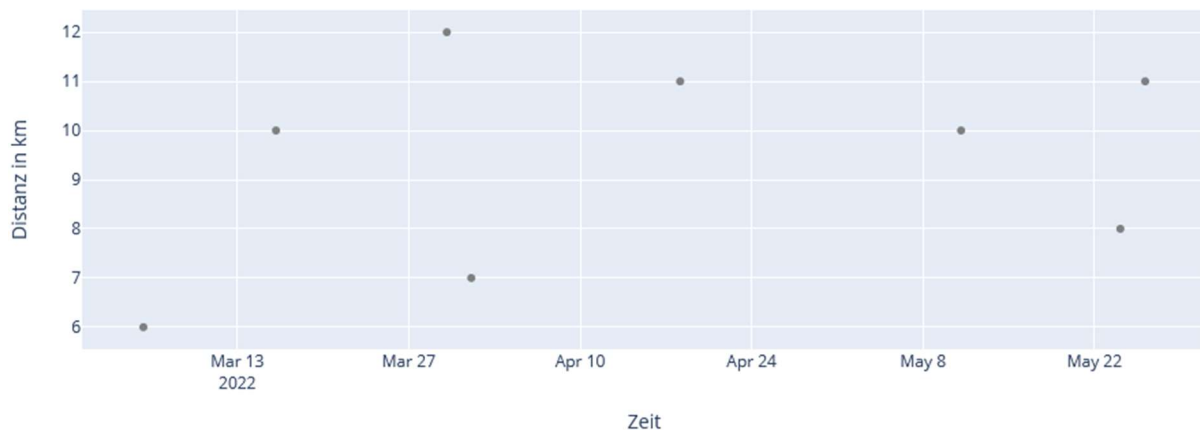


Abbildung 20: Punktdiagramm Visualisierung 2

Für jedes Training wird dann das Punktdiagramm, wie in Abbildung 20 zu sehen, mit dem Datum auf der X-Achse und der Distanz auf der Y-Achse erstellt. Darauf folgt das sogenannte Kategorie-Diagramm, welches die Werte der Kategorien Distanz, durchschnittliche Geschwindigkeit und durchschnittliche Herzfrequenz darstellt. Hierbei werden, wie in Abbildung 21 zu sehen, die zusammengehörigen Daten über die Kategorien hinweg miteinander verbunden.

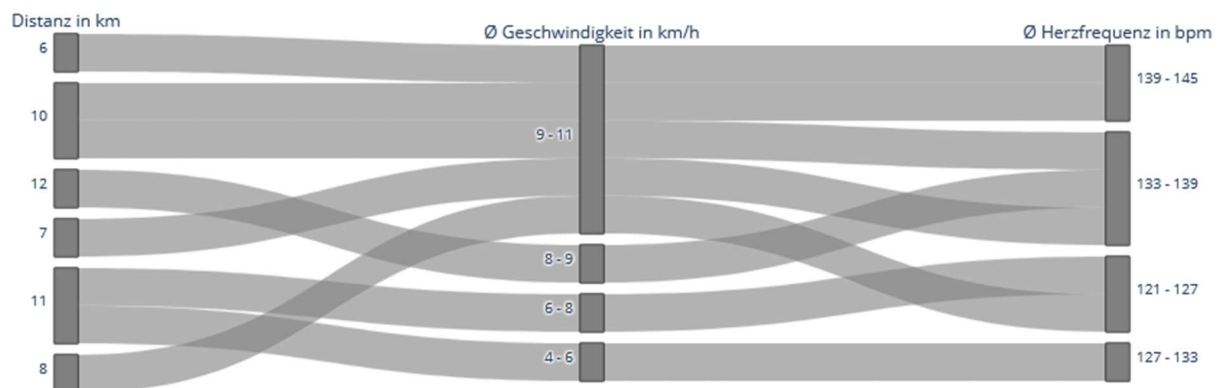


Abbildung 21: Kategorie-Diagramm Visualisierung 2

Die Daten der beiden Diagramme sind miteinander verbunden. Sie bilden ein Verbunds-Diagramm. So können Zusammenhänge hergestellt werden. Implementiert ist dieses Diagramm mit der Plotly Bibliothek. Das Erkennen der Zusammenhänge zwischen beiden Visualisierungen wird durch sogenanntes Brushing & Linking ermöglicht. Der Nutzer markiert, wie in Abbildung 22 zu sehen, zuerst Daten in einer der Visualisierungen. Diese Daten werden dann am Ort der Markierung farblich gekennzeichnet. Zusätzlich werden aber auch die zugehörigen Daten in der anderen Visualisierung markiert. Durch die gleiche, farbliche Markierung wird klar, die Daten gehören zusammen. [28]

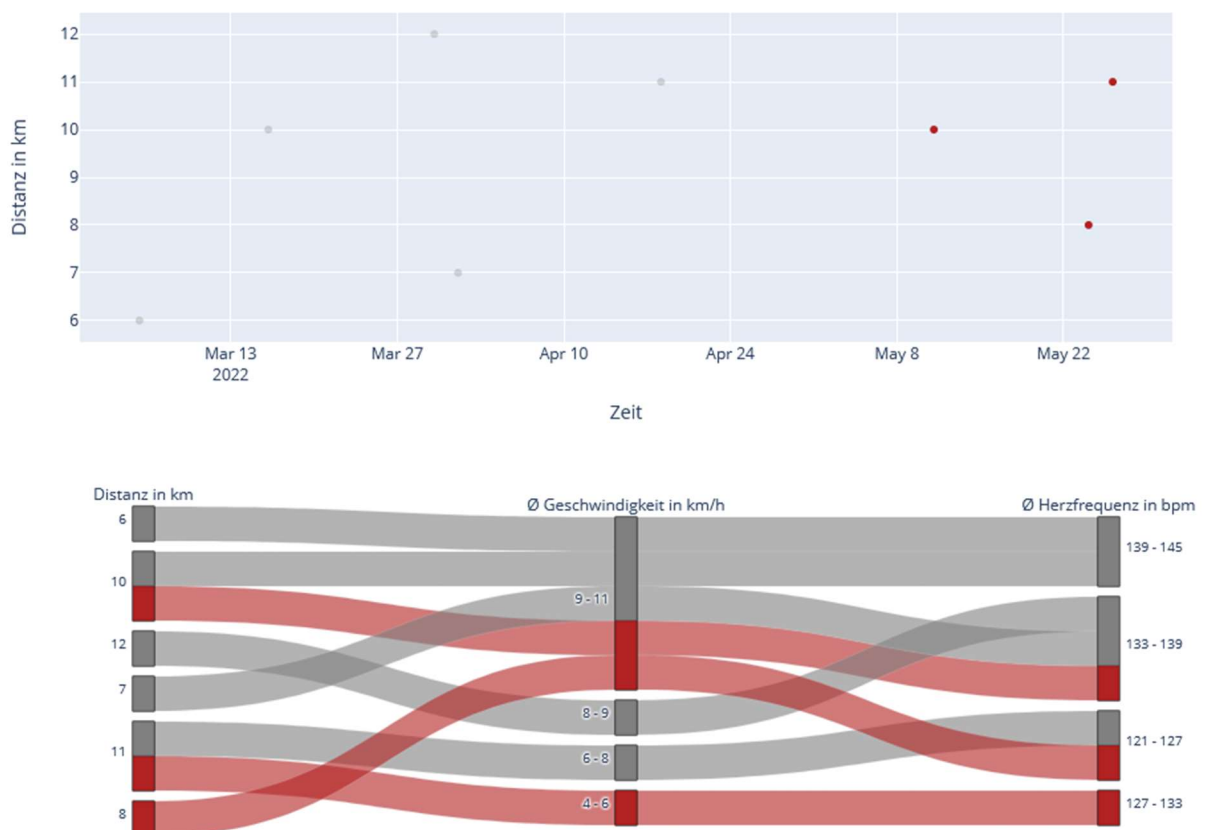


Abbildung 22: Brushing & Linking in Visualisierung 2

Brushing & Linking ist in dieser Visualisierung in beide Richtungen möglich. Über ein Lasso-Werkzeug können mehrere Punkte im Punktdiagramm markiert werden. Über einfaches klicken kann eine Gruppe in einer Kategorie des Kategorie-Diagramms markiert werden. Durch Klicken neben das Diagramm oder durch Auswählen keines Punktes können die Markierungen entfernt werden.

Neben Brushing & Linking kann der Nutzer die Gruppen des Kategorie-Diagramms noch beliebig verschieben und beispielsweise absteigend nach Größe der Gruppe oder aufsteigend nach der durchschnittlichen Herzfrequenz sortieren. In dem Punktdiagramm ist außerdem noch Panning und Zooming möglich. So kann der Nutzer bei einer großen Menge an Daten bestimmte Ausschnitte genauer analysieren.

#### 5.4.3. Herausforderungen

Die größte Herausforderung bei der Implementierung dieser Visualisierung war das Gruppieren der Daten. Das Gruppieren der Daten führt zu dem Verlust von Informationen. Statt die genaue durchschnittliche Geschwindigkeit oder die genaue Distanz zu kennen ist nur noch eine Rundung, bzw. eine Spanne bekannt. Das ist von Nachteil. Jedoch ist Gruppieren dennoch notwendig, da ohne das Gruppieren die Daten unübersichtlich werden und keine Zusammenhänge festzustellen sind.

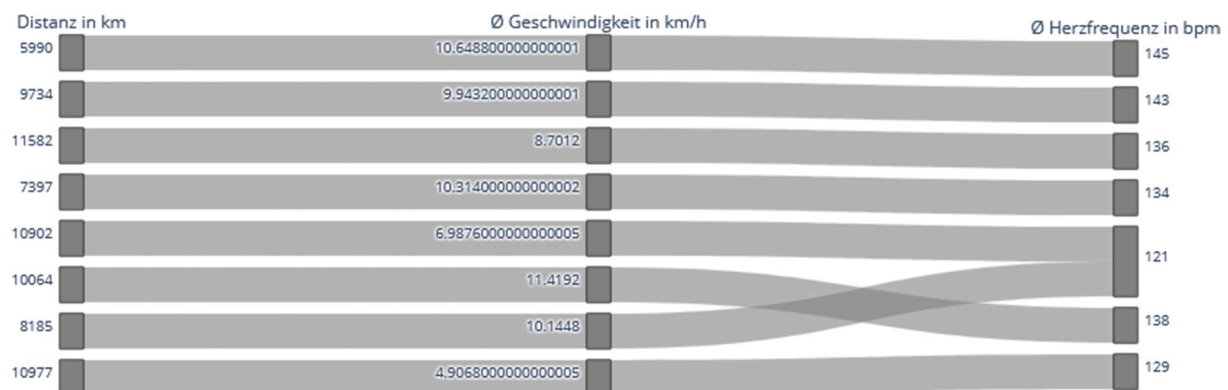


Abbildung 23: Nicht gruppiertes Kategorie-Diagramm

In Abbildung 23 ist ein Kategorie-Diagramm ohne Gruppierung zu sehen. Die Daten sind unübersichtlich und unnötig genau. Ein Kategorie-Diagramm ohne Gruppierung gleicht einer einfachen Tabelle.

Daher werden die Daten gruppiert. Je nach Anzahl und Varianz der Daten sind jedoch eine unterschiedliche Anzahl an Gruppen sinnvoll. Wenn in den Daten beispielsweise schnelle Sprints und langsame Marathons enthalten sind ist die Spanne der durchschnittlichen Geschwindigkeit sehr hoch. Um möglichst viele Informationen zu erhalten sollte dann eine höhere Anzahl an Gruppen gewählt werden. Durch interaktives Ausprobieren und Ändern der Gruppenanzahl kann der Nutzer iterativ ein sinnvoller Wert ermitteln.

#### 5.4.4. Bewertung

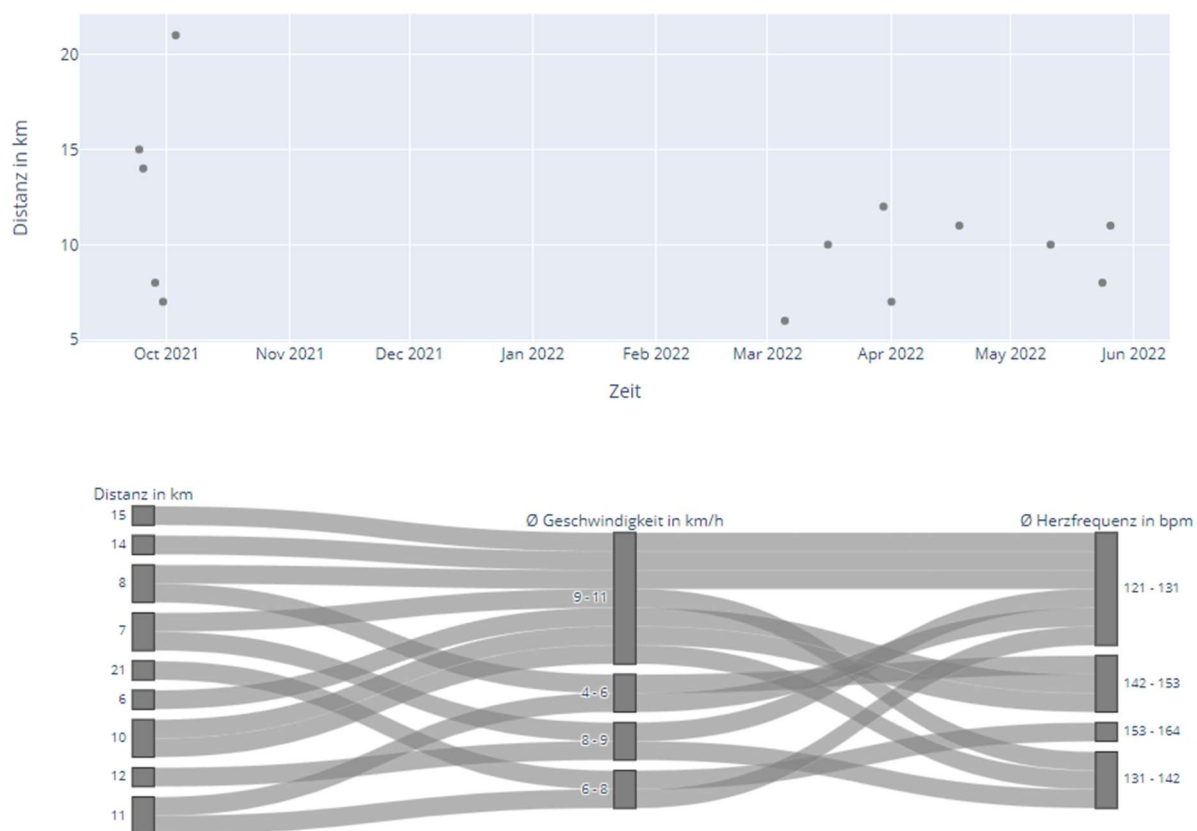


Abbildung 24: Beispiel Visualisierung 2

In Abbildung 24 ist ein Beispiel der Visualisierung mit 13 Daten zu sehen. Anhand Tufte's Prinzipien der Graphical Excellence und Graphical Integrity soll die Visualisierung nun bewertet werden.

Erfüllt die Visualisierung Tufte's Prinzipien der Graphical Excellence [33]?

- Zeigt Sie die Daten?

Ja, Sie zeigt die Daten.

- Regt Sie den Betrachter zum Nachdenken an?

Ja, Sie regt den Betrachter zum Nachdenken an. Der Betrachter sieht die visualisierten Trainings, interagiert mit Ihnen und kann so Muster feststellen. Diese können ihn zum Nachdenken über die vergangenen Trainings anregen.

- Vermeidet Sie es zu verzerren, was die Daten sagen?

Ja, die Visualisierung stellt die Daten unverzerrt da.

- Zeigt Sie viele Zahlen in wenig Raum?

Ja, Sie zeigt in zwei kleinen Diagrammen eine nahezu beliebige Anzahl an Datensätzen, welche Informationen über Distanz, Datum, durchschnittliche Geschwindigkeit und durchschnittliche Herzfrequenz enthalten.

- Bringt Sie große Datensätze in einen Zusammenhang?

Ja, Sie bringt die Datensätze mehrerer Trainings in einen Zusammenhang. Der Nutzer hat die Möglichkeit diese über Interaktionen wie Brushing & Linking zu erkennen.

- Unterstützt Sie Vergleiche zwischen den Daten?

Ja, Sie erlaubt es die Trainings anhand der Distanz, des Datums, der durchschnittlichen Geschwindigkeit und der durchschnittlichen Herzfrequenz miteinander zu vergleichen.

- Präsentiert Sie die Daten in verschiedenen Detailgraden?

Nein, Sie präsentiert die Daten nur auf einem Detailgrad.

- Hat Sie einen klaren Zweck?

Ja, Sie stellt vergangene Trainings dar, ermöglicht es diese besser zu verstehen und so Trends festzustellen.

Erfüllt die Visualisierung Tufte's Prinzipien der Graphical Integrity [33]?

- Entsprechen Visualisierungen von Zahlen Ihren tatsächlichen Proportionen?

Ja, die Visualisierung der Zahlen entspricht Ihren tatsächlichen Proportionen.

- Sind Beschriftungen klar und detailliert?

Ja, die Beschriftungen sind klar und detailliert. Aber möglicherweise ist das Durchschnitt-Symbol  $\bar{x}$  nicht allen bekannt und es sollte stattdessen ausgeschrieben werden.

- Hat die Visualisierung keinen versteckten Zweck, sondern zeigt nur Schwankungen der Daten?

Ja, es werden nur die Schwankungen der Daten gezeigt.

- Werden bekannte Einheiten verwendet?

Ja, es werden die bekannten Einheiten Kilometer, km/h und bpm verwendet.

- Entspricht die Zahl der visualisierten Dimensionen der Zahl der Dimensionen der Daten?

Ja. Sowohl die Daten als auch die Visualisierung ist 4-dimensional.

- Impliziert die Visualisierungen nichts?

Nein. Die Daten implizieren, dass es sich um Fitness-Daten handelt. Im Rahmen der Studienarbeit wird dies aber explizit erwähnt. Ein entsprechender Titel oder eine Beschreibung wäre jedoch zusätzlich angebracht.



#### 5.4.5. Mögliche Weiterentwicklung des Prototyps

Der Prototyp könnte auf mehrere Arten weiterentwickelt werden. Es könnten beispielsweise mehr Informationen zu Verfügung gestellt werden. Neben Geschwindigkeit und Herzfrequenz könnte beispielsweise noch die Höhe des Trainings Teil des Kategorie-Diagramms sein.

Die bei der Bewertung ermittelten Mängel können angegangen werden. So könnte der Visualisierung beispielsweise ein weiterer Detailgrad hinzugefügt werden. Dieser könnte bei Hovern über einen Punkt im Punktdiagramm beispielsweise die genauen Werte für durchschnittliche Geschwindigkeit und durchschnittliche Herzfrequenz darstellen.

### 5.5. Weitere Visualisierungsideen

Neben den zwei implementierten Visualisierungen sind mit den Fitness-Daten noch andere Visualisierungen denkbar. Die hier implementierten Visualisierungen fokussieren sich auf die Analyse mehrerer Trainings und deren Zusammenhänge. Der Nutzer schaut über einen längeren Zeitraum auf seine Trainings zurück und kann so erkennen, ob er seine Ziele erreicht und Fortschritte gemacht hat.

Ein alternativer Ansatz ist die Visualisierung einzelner Trainings. Statt eine Ansammlung von Trainingsdaten zu visualisieren wird nur ein Datensatz ausführlich visualisiert.

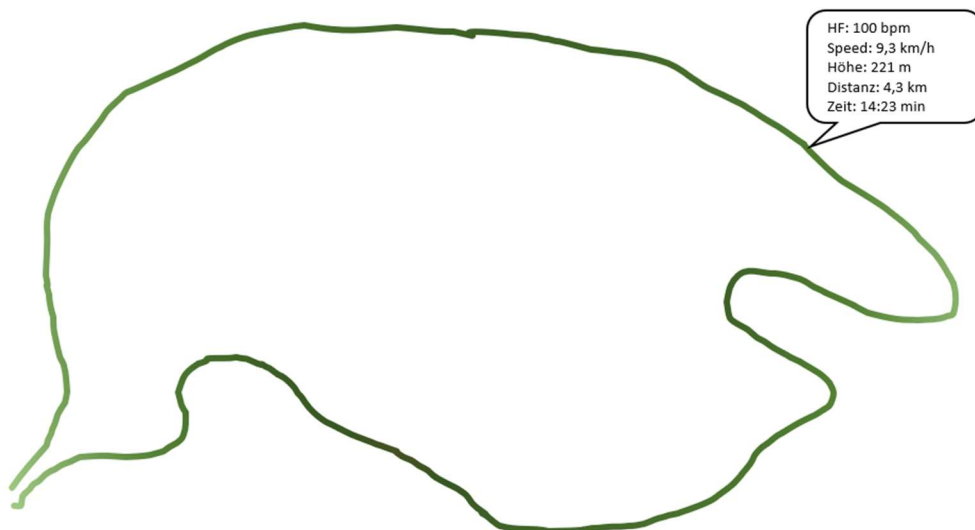


Abbildung 25: Mockup Visualisierung einzelnes Training

In Abbildung 25 ist ein einfaches Mockup einer solchen Visualisierung zu sehen. Wie in Visualisierung 1 ist hier der Streckenverlauf skizziert. Statt jedoch den Streckenverlauf einfarbig einzufärben wird hier der Streckenverlauf nach der Intensität des Trainings zu den konkreten Zeitpunkten gefärbt. So kann auf einen Blick erkannt werden, wann das Training wie intensiv war. Durch das Hovern über den Streckenverlauf könnten weitere Informationen zu dem Trainings-Status zu dem konkreten Zeitpunkt vermittelt werden. Das können Informationen wie Herzfrequenz, Geschwindigkeit, Höhe oder andere sein. Bei Training nach einem Trainingsplan

würde diese Visualisierung ermöglichen, das Training mit dem Plan abzugleichen. Wenn beispielsweise ein Intervalltraining mit Phasen hoher Belastung gefolgt von Phasen niedriger Belastung geplant ist, könnte mit einem Blick auf die Visualisierung erkannt werden, ob dies umgesetzt wurde. Die Phasen könnten weiter noch genauer analysiert werden. Durch das Hovern kann überprüft werden, wie intensiv die Phasen hoher Belastung waren und wie entspannend die Phasen niedriger Belastung.

## 6. Fazit

In dieser Studienarbeit sollte der Visual Analytics Prozess dargestellt und Visualisierungen von Fitness-Daten erstellt werden. Es wurde analysiert welchen Einfluss Fitness-Tracker und die von Ihnen produzierte Daten auf die Gesundheit haben. Sie haben das Potenzial den Nutzer zu motivieren und ihn über Fortschritt oder Stagnation des Trainingserfolgs zu informieren. Das Tracken bestimmter Gesundheitsdaten wie der Herzfrequenz kann Auskunft über einzelne Trainings und deren Intensität geben. Mit dieser Information können Trainingspläne erstellt werden, welche speziell auf den Nutzer zugeschnitten sind.

Weiter wurde der Visual Analytics Prozess vorgestellt. Es wurde auf die Daten eingegangen, welche Grundlage des Prozesses sind. Dann wurde die Visualisierung und die Modellbildung besprochen, welche zentral für den Visual Analytics Prozess sind. Eine wichtige Rolle dabei spielt die Interaktion. Die Interaktion erlaubt es dem Nutzer sein persönliches Expertenwissen einzubringen und so in Kombination mit Visualisierung und Modell neues Wissen zu erlangen. Das neu erlangte Wissen kann wiederum selbst Teil eines neuen Visual Analytics Prozesses sein.

Es wurde außerdem das FIT-Protokoll erläutert, welches verwendet wird, um von Fitness-Trackern stammende Daten zu speichern und zu teilen. Abschließend wurden zusätzlich zu der theoretischen Ausarbeitung zwei, auf Fitness-Daten basierende, Visualisierungs-Prototypen erstellt. Diese Visualisierungen können Nutzern helfen ihr Training besser zu verstehen und so informierte Entscheidungen in der Zukunft zu treffen.

Diese Studienarbeit hat lediglich Visualisierungen der Fitness-Daten erstellt. Weiter kann es sinnvoll sein einen gesamten Visual Analytics Prozess für Fitness-Daten zu entwerfen. Ein solcher Prozess würde zusätzlich die Aufbereitung der Daten, das Erstellen von Modellen und das Wiederverwenden des gewonnenen Wissens enthalten. Solche Prozesse existieren bereits im professionellen Umfeld für beispielsweise Ärzte oder zur Analyse der Gesundheit einer Bevölkerung [39, 40]. Tools zur Analyse von persönlichen Fitness-Daten über einen vollständigen Visual Analytics Prozess sind jedoch noch nicht für Endnutzer zugänglich.

Langfristig ist dies jedoch notwendig. Die Welt produziert immer mehr Daten, welche verstanden werden müssen. Die Nutzung von Fitness-Trackern steigt und die Features der Tracker werden immer umfangreicher. Ohne einen Prozess, der uns diese Daten verstehen lässt, werden sie für den Nutzer wertlos. Es muss möglich werden die Menge an komplexer Daten in einen Zusammenhang zu bringen, welcher dem Nutzer hilft gesünder zu leben. Ein Teil dieses Prozesses wurde in dieser Arbeit umgesetzt. Jedoch ist viel Potenzial für das Verständnis der Daten noch nicht ausgeschöpft.

## Literaturverzeichnis

- [1] B. J. McCarthy, "One in Five U.S. Adults Use Health Apps, Wearable Trackers," *Gallup*, 11 Dec., 2019. <https://news.gallup.com/poll/269096/one-five-adults-health-apps-wearable-trackers.aspx> (accessed: Dec. 19 2021).
- [2] Y. N. Harari, *21 lessons for the 21st century*. New York: Spiegel & Grau, 2018.
- [3] A. Karahanoğlu, R. Gouveia, J. Reenalda, and G. Ludden, "How Are Sports-Trackers Used by Runners? Running-Related Data, Personal Goals, and Self-Tracking in Running," *Sensors*, vol. 21, no. 11, p. 3687, 2021, doi: 10.3390/s21113687.
- [4] J. Gaudet, F. Gallant, and M. Bélanger, "A Bit of Fit: Minimalist Intervention in Adolescents Based on a Physical Activity Tracker," *JMIR mHealth and uHealth*, vol. 5, no. 7, e92, 2017, doi: 10.2196/mhealth.7647.
- [5] D. Leutner, "Perspektiven pädagogischer Interventionsforschung," *Qualitätsentwicklung in der Berufsbildungsforschung*, vol. 2013.
- [6] J. O. Prochaska and W. F. Velicer, "The transtheoretical model of health behavior change," *American journal of health promotion : AJHP*, vol. 12, no. 1, pp. 38–48, 1997, doi: 10.4278/0890-1171-12.1.38.
- [7] Y. Kim, A. Lumpkin, M. Lochbaum, S. Stegemeier, and K. Kitten, "Promoting physical activity using a wearable activity tracker in college students: A cluster randomized controlled trial," *Journal of sports sciences*, vol. 36, no. 16, pp. 1889–1896, 2018, doi: 10.1080/02640414.2018.1423886.
- [8] T. Leskinen *et al.*, "The Effect of Consumer-based Activity Tracker Intervention on Physical Activity among Recent Retirees-An RCT Study," *Medicine and science in sports and exercise*, vol. 53, no. 8, pp. 1756–1765, 2021, doi: 10.1249/MSS.0000000000002627.
- [9] S. S. Martin *et al.*, "mActive: A Randomized Clinical Trial of an Automated mHealth Intervention for Physical Activity Promotion," *JAHA*, vol. 4, no. 11, 2015, doi: 10.1161/JAHA.115.002239.
- [10] W. M. A. Franssen, Franssen, Gregor H. L. M., J. Spaas, F. Solmi, and B. O. Eijnde, "Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials," *Int J Behav Nutr Phys Act*, vol. 17, no. 1, p. 57, 2020, doi: 10.1186/s12966-020-00955-2.
- [11] S. R. Wack, K. A. Crosland, and R. G. Miltenberger, "Using goal setting and feedback to increase weekly running distance," *Journal of applied behavior analysis*, vol. 47, no. 1, pp. 181–185, 2014, doi: 10.1002/jaba.108.
- [12] E. C. Lee, M. S. Fragala, S. A. Kavouras, R. M. Queen, J. L. Pryor, and D. J. Casa, "Biomarkers in Sports and Exercise: Tracking Health, Performance, and Recovery in Athletes," *Journal of strength and conditioning research*, vol. 31, no. 10, pp. 2920–2937, 2017, doi: 10.1519/JSC.0000000000002122.
- [13] S. L. Halson, "Monitoring training load to understand fatigue in athletes," *Sports Med*, 44 Suppl 2, no. 2, S139–47, 2014, doi: 10.1007/s40279-014-0253-z.
- [14] S. A. Rawashdeh, D. A. Rafeldt, and T. L. Uhl, "Wearable IMU for Shoulder Injury Prevention in Overhead Sports," *Sensors*, vol. 16, no. 11, 2016, doi: 10.3390/s16111847.
- [15] A. Eisenhut and F. Zintl, *Ausdauertraining: Grundlagen Methoden Trainingssteuerung*. München: BLV Buchverlag, 2014.

- [16] G. Neumann, A. Pfützner, and A. Berbalk, *Optimiertes Ausdauertraining*, 7th ed. Aachen: Meyer & Meyer, 2012.
- [17] A. Middelweerd, J. S. Mollee, C. N. van der Wal, J. Brug, and S. J. Te Velde, "Apps to promote physical activity among adults: a review and content analysis," *Int J Behav Nutr Phys Act*, vol. 11, p. 97, 2014, doi: 10.1186/s12966-014-0097-9.
- [18] X. Gui, Y. Chen, C. Caldeira, D. Xiao, and Y. Chen, "When Fitness Meets Social Networks," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver Colorado USA, 05022017, pp. 1647–1659.
- [19] Y. Chen and P. Pu, "HealthyTogether," in *Proceedings of the Second International Symposium of Chinese CHI on - Chinese CHI '14*, Toronto, Ontario, Canada, 2014, pp. 25–34.
- [20] M. Budig, V. Höltnke, and M. Keiner, "Accuracy of optical heart rate measurement and distance measurement of a fitness tracker and their consequential use in sports," *Ger J Exerc Sport Res*, vol. 49, no. 4, pp. 402–409, 2019, doi: 10.1007/s12662-019-00621-1.
- [21] H. Weghorn, "Unsubstantial Health and Sports Monitoring Reliability of Commercial Fitness Tracker Bracelets Induced by Their All-in-One Sensing Unit Approach - Experimental Evaluation of Measurement Accuracy in Dynamic and in Steady Physical Effort Scenarios," in *Communications in Computer and Information Science, Sport Science Research and Technology Support*, J. Cabri, P. Pazarat-Correia, and J. Vilas-Boas, Eds., Cham: Springer International Publishing, 2019, pp. 55–74.
- [22] J. Lee and J. Finkelstein, "Consumer sleep tracking devices: a critical review," *Studies in health technology and informatics*, vol. 210, pp. 458–460, 2015.
- [23] A. Papic and K. Durdevic, "Young Adults' Privacy Concerns about Wearable Technology," in *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*, Opatija, Croatia, Sep. 2021 - Oct. 2021, pp. 537–541.
- [24] P. C. Wong and J. Thomas, "Visual analytics," *IEEE computer graphics and applications*, vol. 24, no. 5, pp. 20–21, 2004, doi: 10.1109/mcg.2004.39.
- [25] D. A. Keim, F. Mansmann, A. Stoffel, and H. Ziegler, "Visual Analytics," in *Encyclopedia of Database Systems*, L. LIU and M. T. ÖZSU, Eds., Boston, MA: Springer US, 2009, pp. 3341–3346.
- [26] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler, "Visual Analytics: Scope and Challenges," in *Lecture notes in computer science, State-of-the-art survey 0302-9743*, vol. 4404, *Visual data mining: Theory, techniques and tools for visual analytics / Simeon J. Simoff, Michael H. Böhlen, Arturas Mazeika (eds.)*, S. J. Simoff, M. H. Böhlen, and A. Mazeika, Eds., Berlin: Springer, 2008, pp. 76–90.
- [27] R. J. Andrews, *Info we trust: How to inspire the world with data*. Hoboken New Jersey: John Wiley & Sons, 2019.
- [28] J. Kohlhammer, *Visual business analytics: Effektiver zugang zu Daten und Informationen*. Heidelberg, Germany: dpunkt.verlag, 2013. [Online]. Available: <http://site.ebrary.com/lib/alltitles/Doc?id=10821568>
- [29] T. A. Team, "How, When, and Why Should You Normalize / Standardize / Rescale Your Data?," *Towards AI*, 29 May., 2020. <https://towardsai.net/p/data-science/how-when-and-why-should-you-normalize-standardize-rescale-your-data-3f083def38ff> (accessed: May 25 2022).
- [30] F. E. Grubbs, "Sample Criteria for Testing Outlying Observations," *Ann. Math. Statist.*, vol. 21, no. 1, pp. 27–58, 1950, doi: 10.1214/aoms/1177729885.

- [31] Z. He, X. Xu, and S. Deng, “Discovering cluster-based local outliers,” *Pattern Recognition Letters*, vol. 24, 9-10, pp. 1641–1650, 2003, doi: 10.1016/S0167-8655(03)00003-5.
- [32] S. Hawkins, H. He, G. Williams, and R. Baxter, “Outlier Detection Using Replicator Neural Networks,” in 2002, pp. 170–180. [Online]. Available: [https://link.springer.com/chapter/10.1007/3-540-46145-0\\_17](https://link.springer.com/chapter/10.1007/3-540-46145-0_17)
- [33] E. R. Tufte, *The visual display of quantitative information*, 2nd ed. Cheshire Conn.: Graphics Press, 2001.
- [34] W. Krämer, *So lügt man mit Statistik*, 1st ed. Frankfurt am Main: Campus Verlag, 2015. [Online]. Available: [http://ebooks.ciando.com/book/index.cfm/bok\\_id/1984045](http://ebooks.ciando.com/book/index.cfm/bok_id/1984045)
- [35] P. Saraf, R. R Sedamkar, and S. Rathi, “PrefixSpan Algorithm for Finding Sequential Pattern with Various Constraints,” 2249-0868, vol. 9, no. 3, pp. 37–41, 2015, doi: 10.5120/ijais15-451380.
- [36] B. Shneiderman, “The eyes have it: a task by data type taxonomy for information visualizations,” in *Proceedings 1996 IEEE Symposium on Visual Languages*, Boulder, CO, USA, Sep. 1996, pp. 336–343.
- [37] *FIT Protocol | FIT SDK | Garmin Developers* (accessed: Nov. 15 2021).
- [38] *Choosing Colormaps in Matplotlib — Matplotlib 3.5.0 documentation*. [Online]. Available: <https://matplotlib.org/3.5.0/tutorials/colors/colormaps.html> (accessed: Jun. 10 2022).
- [39] Y. Zhao *et al.*, “Visual Analytics for Health Monitoring and Risk Management in CARRE,” in *LNCS sublibrary: SL3 - Information systems and applications, incl. Internet/Web, and HCI*, vol. 9654, *E-learning and games: 10th International Conference, Edutainment 2016, Hangzhou, China, April 14-16, 2016 : revised selected papers / Abdennour El Rhalibi, Feng Tian, Zhigeng Pan, Baoquan Liu (eds.), A. El Rhalibi, Ed., Cham?: Springer, 2016*, pp. 380–391.
- [40] O. R. Ferch, “Visual analytics of geolocated open fitness data of children and teens,” 2021.