

MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG

Projektbericht zum Modul Information Retrieval und Visualisierung Sommersemester  
2021

## **Visualisierung von Daten des Videospiels Fifa 21**

Eingereicht bei: Dr. Alexander Hinneburg

Eingereicht von: Johannes Lange

Git-Repository: [https://github.com/JohannesLange/Visualisierung\\_FIFA19](https://github.com/JohannesLange/Visualisierung_FIFA19)

Eingereicht am: 10. September 2021

# Inhaltsverzeichnis

<b>Abkürzungsverzeichnis</b>	<b>III</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Anwendungshintergrund . . . . .	2
1.2 Zielgruppen . . . . .	2
1.3 Überblick und Beiträge . . . . .	4
<b>2 Daten</b>	<b>4</b>
2.1 Technische Bereitstellung der Daten . . . . .	5
2.2 Datenvorverarbeitung . . . . .	5
<b>3 Visualisierungen</b>	<b>6</b>
3.1 Analyse der Anwendungsaufgaben . . . . .	6
3.2 Anforderungen an die Visualisierungen . . . . .	6
3.3 Präsentation der Visualisierungen . . . . .	7
3.3.1 Visualisierung Eins . . . . .	7
3.3.2 Visualisierung Zwei . . . . .	8
3.3.3 Visualisierung Drei . . . . .	8
3.4 Interaktion . . . . .	9
<b>4 Implementierung</b>	<b>10</b>
<b>5 Anwendungsfälle</b>	<b>10</b>
5.1 Anwendung Visualisierung Eins . . . . .	11
5.2 Anwendung Visualisierung Zwei . . . . .	12
5.3 Anwendung Visualisierung Drei . . . . .	13
<b>6 Verwandte Arbeiten</b>	<b>13</b>
<b>7 Zusammenfassung und Ausblick</b>	<b>15</b>

## Abbildungsverzeichnis

1	Darstellung einer FUT Karte inklusive relevanter Werte am Beispiel Cristiano Ronaldo Quelle: <a href="https://www.ea.com/de-de/games/fifa/fifa-21/news/fifa-21-player-ratings-best-strikers-st-cf">https://www.ea.com/de-de/games/fifa/fifa-21/news/fifa-21-player-ratings-best-strikers-st-cf</a> . . . . .	3
2	Darstellung des Scatterplots Quelle: eigene Darstellung . . . . .	8
3	Darstellung der Parallelen Koordinaten Quelle: Eigene Darstellung . . . . .	9
4	Ausschnitt aus Baumdarstellung Quelle: Eigene Darstellung . . . . .	9
5	Darstellung des Scatterplot und dessen <i>on:hover</i> Funktionalität Quelle: eigene Darstellung . . . . .	11
6	Darstellung der Parallelen Koordinaten inklusive <i>on:hover</i> Funktionalität Quelle: eigene Darstellung . . . . .	12
7	Darstellung der Parallelen Koordinaten inklusive <i>on:hover</i> Funktionalität Quelle: eigene Darstellung . . . . .	13
8	Darstellung des Baumdiagramms Quelle: eigene Darstellung . . . . .	14

## Abkürzungsverzeichnis

**CB** Centerback

**FUT** Fifa Ultimate Team

**RB** Rightback

# 1 Einleitung

Die Visualisierung von Daten nimmt mit Hinblick auf Big Data und damit immer unübersichtlicheren Datengrundlagen an Bedeutung zu. Um aus großen Mengen von Daten neue Informationen zu gewinnen reicht es nicht aus die Daten direkt zu analysieren. Durch die Anwendung der richtigen Visualisierungstechniken können neue Informationen gewonnen werden.

Besonders Fans von kompetitiven Videospielen versuchen oft jeden möglichen Vorteil zu nutzen um gewinnen zu können. Das Spiel FIFA 21, dessen Daten in dieser Arbeit visualisiert werden, bildet dabei keine Ausnahme. Früher wurden dazu häufig Guides verwendet in denen hunderte Seiten Tipps und Tricks abgedruckt worden sind. Diese Guides bieten jedoch keine Interaktion und sind oft aus der subjektiven Sicht eines Autors verfasst. Moderne Visualisierungstechniken erlauben es jedoch deutlich mehr Informationen zu extrahieren und für die Spieler interaktiv nutzbar zu machen. Der gegebene Datensatz ist äußerst umfangreich, zu über 16000 Sportlern sind jeweils 107 Variablen festgehalten. Diese reichen von *Name* über *Verein* bis hin zu *Schusskraft* eines Spielers. Zwar sind diese Daten dem Videospiel FIFA 21 entnommen, Daten wie Name, Alter, Verein oder ähnliche beruhen aber auf echten Daten der Spieler. Weiterhin versucht der Hersteller des Spiels auch die restlichen Daten so nah an der Realität zu verankern wie möglich. So besteht die Hoffnung mittels Visualisierungstechniken noch weitgreifendere Erkenntnisse aus dem gegebenen Datensatz ziehen zu können, als solche die Spielern des Spiels dienen.

Spieler des Spiels möchten möglicherweise herausfinden, welcher Fussballer ihr Team verbessert aber wenig kostet also: *Welche Spieler hat das höchste Overall bei geringstem Gehalt?*

Eine Weitere Frage könnte sich beim Bau eines Teams bestehend aus Bundesliga Teams ergeben wenn noch ein Spieler einer bestimmten Position benötigt wird: *Welche Linksverteidiger spielen in der Bundesliga?*

Außerdem könnten sich Fragen nach den besten Spielern bezüglich individueller Kriterien ergeben: *Welcher Stürmer ist schnell, Schussstark, kann gut dribbeln und ist groß?*

Diese konkreten Fragen beziehen sich alle auf Fragen welche Anforderungen von Spielern des Spiels darstellen.

Ob Fragen von anderen Anforderungsgruppen, wie bspw. Sportwissenschaftlern beantwortet werden können: *Wann erreichen Fussballer ihren Peak? Sind größere Fussballer besser, schneller, etc.?* ist dabei nicht klar, da alle nicht erhobenen Daten von Mitarbeitern des Herstellers festgelegt worden sind und somit nicht genau den tatsächlichen Daten entsprechen können. Trotzdem lohnt es sich dies zu überprüfen und mit anderen wissenschaftlichen Arbeiten abzugleichen.

## 1.1 Anwendungshintergrund

Im folgenden werden die drei verwendeten Visualisierungstechniken kurz vorgestellt und erklärt. Bei der ersten Visualisierungstechnik handelt es sich um einen Scatterplot.[1] Mit diesem lassen sich Zusammenhänge zwischen zwei verschiedenen, numerischen Attributen der Fussballer genauer untersuchen. So kann bspw. gezeigt werden ob sich das Alter von Fussballern auf deren Bewertung im Spiel auswirkt.

Die zweite Visualisierungstechnik ist die der parallelen Koordinaten.[2] Diese eignet sich aufgrund ihrer Beschaffenheit dazu mehrdimensionale Daten auf übergreifende Trends zu untersuchen. Trotzdem können sie auch genutzt werden um spezifische Fussballer konkret zu untersuchen und deren Werte im Vergleich zu den übergreifenden Daten zu betrachten. Diese Darstellungsform ermöglicht den Vergleich vier verschiedener Attribute der Fussballer. Dabei ist allerdings zu beachten, dass Zusammenhänge innerhalb der Daten nur im Vergleich mit anderen Daten zu erkennen sind. Enthält der Datensatz bspw. nur einen Spieler zeigen die parallelen Koordinaten nur eine gerade Linie an. Weiterhin ist zu beachten, dass lediglich Zusammenhänge zwischen benachbarten Attributen zu erkennen sind. Dies kann bei ungeschulten Anwendern schnell zu Fehlinterpretationen führen. In diesem Fall ist sich für parallele Koordinaten mit vier verschiedenen Attributen entschieden worden. Dadurch können genügend Informationen angezeigt werden ohne eine gewisse Übersichtlichkeit zu verlieren. Diese Visualisierungsform kann besonders dazu genutzt werden mehrere Attribute eines Fussballers gleichzeitig zu untersuchen. Damit kann der Frage nach der Eignung eines Fussballers zum Einsatz auf einer spezifischen Position nachgegangen werden.

Bei der dritten Visualisierungstechnik handelt es sich um die explizite Baumdarstellung. Durch sie kann ein Überblick über hierarchische Strukturen ermöglicht werden. In diesem Fall ist dies die Fussballliga als oberstes Element, darauf folgen die Vereine der Liga und zum Schluss die Fussballer der Vereine. Von den drei Techniken ist diese aufgrund ihrer Bekanntheit vermutlich am intuitivsten zu verstehen.

Die Frage, welche sich bei diesen Daten stellt ist, welchen Nutzen sie außerhalb eines Videospiels haben. Deswegen ist mein Ansatz zu vergleichen wie genau diese Daten die Wirklichkeit widerspiegeln. Ein Beispiel dafür könnte sein ob sich das Rating eines Spielers mit der körperlichen Entwicklung eines Spielers bewegt. Sollte dies zutreffen müssten Spieler im Alter ihres physischen Peaks das höchste Rating haben.

## 1.2 Zielgruppen

Da es sich bei den Daten um die eines Videospiels handelt ist davon auszugehen, dass eine der Zielgruppen der Daten Spieler des Videospiels sind. Diese sind weiterhin in die des Spielmodus *Karriere* und die des Spielmodus *Fifa Ultimate Team FUT* zu unterteilen.

Spieler des FUT Modus sammeln die Fussballer wie in einer Art Sammelkarten Spiel. Mit den

Karten der Fussballer (Siehe Abbildung 1) kann dann Online gegen andere Spieler angetreten werden. Dabei sind die aktuellen Werte der Spieler (Siehe sechs Werte auf Abbildung 1) von Relevanz. Da dieser Modus sehr kompetitiv ist kann ein Fussballer bereits durch einen schlechten Wert in einer Kategorie für Spieler unbrauchbar werden. Die Werte eines Spielers können im Spiel nicht verändert werden, deswegen kann so eine Filterung von Spielern vorgenommen werden. Dadurch, dass *FUT* zu den Onlinespielen gehört in denen viele Spieler sich durch *Min-Maxing*[4] einen Vorteil verschaffen wollen ist bei den Anwendern von einem hohen Vorwissen bezüglich der Daten auszugehen. Auf die Visualisierungstechnik bezogen ist nicht davon auszugehen, dass die Funktionsweise der parallelen Koordinaten bekannt ist. Trotzdem bietet sich für diese Zielgruppe die Visualisierung mittels paralleler Koordinaten an, da diese Gruppe genug Interesse aufweist um bereit sein Zeit in das Verständnis einer Visualisierungstechnik zu investieren. Eine weitere für diese Gruppe relevante Visualisierungstechnik ist die explizite Baumdarstellung, durch diese ist es möglich zu untersuchen welche Spieler in welcher Liga und welchem Team spielen so können Spieler dabei unterstützt werden ihr Team passend zusammen zu stellen. Das Verständnis eines Baumdiagramms sollte dabei den meisten intuitiv möglich sein.

Spieler des Karrieremodus spielen das Spiel als Manager eines Vereins. In diesem Spielmodus ist die Entwicklung von jungen Spielern wichtig. Vor allem der Vergleich von Variablen wie Gehalt, Alter, Rating oder Potential ist dabei relevant. In diesem Fall bietet sich eine Darstellung per Scatterplot an, da sich so schnell die aussichtsreichsten Fussballer erkennen lassen. Da der Karrieremodus nicht Online gespielt wird und die Schwierigkeit manuell festgelegt werden kann, ist davon auszugehen, dass das Vorwissen der Gruppe stark schwankt. Während einige Spieler bereits viel Wissen und sehr genaue Informationen benötigen um auf dem höchsten Schwierigkeitsgrad bestehen zu können, benötigen andere erst einmal Grundwissen bezüglich der Spieler. Aus diesem Grund bietet sich auch die Visualisierung mittels eines Baumdiagramms an. Durch diese kann sich ein grober Überblick über Spieler einer Liga gemacht und durch die Möglichkeit des Filterns nach Positionen können grob auch Spieler gescouted werden. Aufgrund des geringen Bedarfs an besonders spezifischen Informationen im Vergleich zum FUT Modus ist nicht mit genügend Ansporn zum Erlernen der Funktionsweise von parallelen Koordinaten zu rechnen. Die Funktionsweise eines Scatterplots sowie eines Baumdiagramms sollte aber den meisten Mitgliedern dieser Zielgruppe intuitiv bekannt sein.



Bei der dritten Zielgruppe, den Sportwissenschaftlern ist von genügend Vorwissen bezüglich der Daten sowie aller Visualisierungstechniken auszugehen. Bei ihnen stellt sich allerdings noch immer die Frage, welchen Mehrwert ihnen die Visualisierungstechniken bieten können. Beim Baumdiagramm ist dabei von keinem Nutzen auszugehen. Scatterplot

Abbildung 1: Darstellung einer FUT Karte inklusive relevanter Werte am Beispiel Cristiano Ronaldo  
Quelle: <https://www.ea.com/de-de/games/fifa/fifa-21/news/fifa-21-player-ratings-best-strikers-st->

und parallele Koordinaten könnten, nach Analyse der Daten auf Realitätsnähe, dazu genutzt werden mögliche Fragestellungen zu beantworten. Weiterhin ist zu sagen, dass dieser Zielgruppe vermutlich nur Erkenntnisse über Trends innerhalb des Datensatzes behilflich sein werden, nicht aber Informationen über spezifische Fussballer.

### 1.3 Überblick und Beiträge

Wie bereits beschrieben sind Guides zu Videospielen bisher meist in Papierform veröffentlicht worden und sind wenig interaktiv. Da das Informationsbedürfnis von Spielern gerade in den letzten Jahren stark zugenommen hat, bietet sich der Einsatz von Visualisierungstechniken auf Basis von Daten, welche das Spiel selbst liefert an. So ist es möglich den gesteigerten Informationsbedarf zu decken ohne direkt Programmierkenntnisse beim Spieler vorauszusetzen. Wie genau die Visualisierungen umgesetzt worden sind und welchen Mehrwert jede einzelne Technik gegenüber anderen Visualisierungsformen bietet ist in Kapitel 3 diskutiert. Welcher Mehrwert sich für spezifische Nutzer ergibt ist in Kapitel 5 beschrieben.

## 2 Daten

Grundsätzlich eignen sich die Daten gut um die gewünschten Fragestellungen beantworten zu können, jedoch enthält der Grunddatensatz einige Felder, die für die Visualisierung nicht nötig sind, deswegen wurde der Datensatz in der Vorvorarbeitung noch verkleinert (Siehe 2.2). Außerdem ist der Datensatz sehr groß, was gerade bei Scatterplot und parallelen Koordinaten zu Unübersichtlichkeit führen kann wenn der ganze Datensatz angezeigt wird, deswegen wurde sich bei diesen Visualisierungen dazu entschieden nach zusätzliche Filterfunktionalitäten einzubauen um dies so übersichtlicher zu gestalten. Da Datenwerte wie Größe, Alter und Rating der Spieler diskret sind, ist entschieden worden die Opazität der Punkte zu verringern, so ist zu erkennen an welchen Stellen sich mehrere Spieler überlagern. Da ab einer bestimmten Anzahl an Spielern die maximale Opazität erreicht ist und eine geringere Opazität pro Punkt einzelne Punkte fast unsichtbar machen würde, wird im Scatterplot beim hovern über einem Punkt angezeigt wie viele Spieler die genau gleichen X und Y Werte aufweisen.

## 2.1 Technische Bereitstellung der Daten

Die Daten sind in einem öffentlichen Github Repository gehostet<sup>1</sup>. Dort liegen die Daten der drei Visualisierungen jeweils als *.csv* vor. Die verwendeten Variablen sowie die Anzahl an Fussballern wurden jeweils der Visualisierungstechnik entsprechend angepasst um den Datensatz so klein wie möglich zu halten und so die Geschwindigkeit der jeweiligen Anwendung zu optimieren. Auch die Daten der Visualisierung als Baumdiagramm liegen als *.csv* vor. Sie werden erst im Programm zu einer *.json* encoded und anschließend wieder zu einem Baumdiagramm decoded. So liegen die Daten nicht starr als *.json* vor. Dies erlaubt das vorfiltern der Daten innerhalb des Programms. Da der Datensatz keine Daten bezüglich der Fussballligen, in denen die Vereine spielen, enthält, sind diese manuell eingefügt worden.

## 2.2 Datenvorverarbeitung

Der Datensatz besteht aus Daten von ca. 17000 einzelnen Fussballern, jeder dieser hat 107 einzelne Variablen. Da diese nicht alle relevant für die Visualisierungen sind und ein kleinerer Datensatz die Visualisierung beschleunigt, sind nur die relevanten Variablen ausgewählt worden. Weiterhin unterscheiden sich die Variablen auch zwischen den verschiedenen Visualisierungstechniken. Aus diesem Grund existieren verschieden gefilterte Datensätze für jede Visualisierung.

Die Vorverarbeitung der Daten ist in *R* durchgeführt worden. Das genutzte Skript (*RData-Preprocessing.r*) liegt im Github vor. Im ersten Schritt des preprocessings sind *NA* Werte aus dem Datensatz entfernt worden. Außerdem ist der Wert *Height* von Fuß und Inch in Zentimeter umgewandelt worden um so eine Schreibweise ohne nicht-numerische Zeichen zu erhalten (Bsp.: 6"2 wird zu 188). Weiterhin ist beim Wert *Weight* die Beschriftung *lbs* entfernt worden um den Wert so als *Integer* für den csv-Decoder von elm erkennbar zu machen. Gleiches ist für die Werte *Wage* und *Value* getan worden. Bei diesen Werten ist die Umwandlung allerdings komplexer, da die Zahlen mit *M* für Millionen und *K* für Tausend abgekürzt sind. Diese Werte sind in Tausend im finalen Datensatz enthalten. Werte mit *M* sind also mit 1000 multipliziert worden, Werte mit *K* sind entsprechend unverändert geblieben und Werte ohne Buchstaben sind durch 1000 geteilt worden (Bsp.: 1.1*M* ist zu 1100 geworden, 90*K* zu 90 und 900 zu 0.9). Nach dem anpassen dieser Werte können die für *Scatterplot* und *Parallele Koordinaten* relevanten Werte in einem neuen *Data.frame* gespeichert und als *.csv* wieder exportiert werden.

Der vorhandene Datensatz ist aus zwei verschiedenen Gründen gekürzt worden, zum einen ist die Darstellung durch die Größe des Datensatzes erheblich verlangsamt worden und zum anderen enthält der Datensatz Fussballer für die nicht alle Werte enthalten sind.

Für die Darstellung des Baumdiagramms wurde sich auf die vier größten europäischen Fussballligen (Premier League, Bundesliga, La Liga und Serie A) konzentriert. Deswegen enthält der Datensatz für diese Visualisierungstechnik nur Daten von Spielern, deren Team Teil einer dieser Ligen ist.

---

<sup>1</sup>[https://github.com/JohannesLange/Visualisierung\\_FIFA19/tree/master](https://github.com/JohannesLange/Visualisierung_FIFA19/tree/master)



Im Git Repository sind unter dem Ordner *data* der ursprüngliche unveränderte Datensatz (*fifa21\_male2.csv*) sowie die veränderten Datensätze (.....) zu finden.

### 3 Visualisierungen

Im folgenden Kapitel werden die drei Visualisierungen vorgestellt. Dabei wird sowohl auf die jeweiligen Einsatzmöglichkeiten der Techniken, als auch auf Anforderungen, die sich aus den Zielstellungen ergeben, eingegangen. Weiterhin werden im weiteren Verlauf die Visualisierungen vorgestellt und Interaktionsmöglichkeiten beschrieben.

#### 3.1 Analyse der Anwendungsaufgaben

Die Visualisierungen sollen Nutzer dabei unterstützen neue Erkenntnisse aus Daten zu ziehen die ihnen ohne die Visualisierung, wenn überhaupt, nur als Liste vorliegen. Dabei bieten Scatterplot und Parallele Koordinaten die Möglichkeit größere Trends innerhalb der Daten zu erkennen: *Wann erreichen Fussballer ihren Peak? Sind größere Fussballer besser, schneller, etc.?* Durch die eingebaute Funktion, dass ein Pfad eines Spielers beim hovern die Farbe ändert und die genauen Werte des Spielers angezeigt werden kann auch ein spezieller Spieler aus der gesamten Liste gefiltert werden. So kann untersucht werden ob dessen Eigenschaften von den generellen Trends abweicht. So kann beispielsweise mittels paralleler Koordinaten ein besonders "guter" Stürmer gefunden werden. Ähnliches bewirkt die Funktion der Farbänderung und Namensanzeige beim hovern über Punkten im Scatterplot.

##### Schreiben über Baumdiagramm

Insgesamt ist zu beachten, dass die Zielgruppen der Visualisierung mit Scatterplots und Baumdiagrammen vermutlich bereits in Kontakt gekommen sind und sie diese deswegen schon interpretieren können. Parallele Koordinaten hingegen sind eine vergleichsweise unbekanntere Visualisierungstechnik. Aus diesem Grund würde es sich anbieten zu den parallelen Koordinaten eine Erklärung der Funktionsweise beizulegen um den Nutzern zu helfen diese Technik zu ihrem vollen Potential nutzen zu können.

**Wie hilft Scatterplot/Parallele Koordinaten/Baumdiagramm die genannten Problemstellungen zu beantworten?**

#### 3.2 Anforderungen an die Visualisierungen

Da die Visualisierungen nicht zum Selbstzweck erstellt werden sondern spezifische Ziele adressieren sollen, müssen sie auch deren Anforderungen erfüllen. Deswegen beschreibt dieser Abschnitt Designentscheidungen der Visualisierungen, die sich aus Anforderungen der Zielprobleme ergeben.

Da es sich bei den Zielgruppen nicht um Programmierer handelt ist eine Anforderung an alle Visualisierungen eine verständliche Darstellung. Dazu gehört die verständliche Beschriftung

der Achsen sowie passende farbliche Kodierung einiger Elemente. Ein weiterer Teil dieser Anforderung ist Übersichtlichkeit zu wahren. Im Fall großer Datenmengen können Visualisierungen schnell unübersichtlich werden und so ihren Nutzen einbüßen. Aus diesem Grund sind in alle drei Visualisierungen verschiedene Filterfunktionen implementiert worden. Dies befriedigt ein weiteres Ziel: nur bestimmte Teilgruppen miteinander vergleichen zu können.

Eine weitere Designentscheidung ergibt sich aus dem Ziel spezifische Fussballer untersuchen zu können. Aus diesem Grund ist sowohl im Scatterplot als auch bei den Parallelen Koordinaten eine *on: hover* Funktion implementiert worden, die den gewählten Punkt oder Pfad hervorhebt und Name sowie Informationen zu dem jeweiligen Spieler anzeigt. Des Weiteren Um den Nutzern einen Mehrwert gegenüber analogen Visualisierungen zu bieten und den Bedarf selbstgewählter Vergleiche von Daten zu decken, ist in die Visualisierung des Scatterplots und der Parallelen Koordinaten die Möglichkeit eingebaut worden die zu vergleichenden Attribute selbst zu wählen.

Beim Implementieren der Visualisierungen müssen aufgrund der Anforderungen der Ziele einige Anpassungen im Vergleich zu den Grundformen dieser Darstellung, welche in den Übungen erstellt worden sind, vorgenommen werden. Besonders die Größe des Datensatzes sorgt dafür, dass viele Trends innerhalb der Daten kaum visuell zu erkennen sind. Aus diesem Grund sind in allen drei Visualisierungsformen verschiedene Filterfunktionen implementiert worden. So werden neben, den kompletten Datensatz übergreifenden Analysen, auch spezifische Untersuchungen an Hand Spieler einzelner Länder, Vereine und/oder Positionen möglich. Weiterhin ist wie bereits beschrieben in der Visualisierungsform der Parallelen Koordinaten eine Anzeige der genauen Werte eines Spielers, beim hovern über dessen Pfad, implementiert worden um überhaupt individuelle Spieler aus dem Datensatz erkennbar zu machen. Dies ist ebenfalls im Scatterplot implementiert worden. Das Baumdiagramm ist hingegen lediglich mit einem Filter ausgestattet um zu verhindern, dass der Nutzen der Visualisierungsform aufgrund von Unübersichtlichkeit, die bei breiten Baumdiagrammen schnell eintritt, abhanden kommt.

### **3.3 Präsentation der Visualisierungen**

Im folgenden Kapitel werden die drei Visualisierungsformen vorgestellt. Weiterhin wird auf Funktionalität und Interaktionsmöglichkeiten genauer eingegangen.

#### **3.3.1 Visualisierung Eins**

Bei der ersten Visualisierung handelt es sich um einen Scatterplot. Mit diesem können zwei Attribute der Fussballer miteinander verglichen werden. Welche diese sind kann vom Anwender per Knopfdruck festgelegt werden. Im Scatterplot wird jeder Spieler als ein Punkt dargestellt. Da im Datensatz viele diskrete Werte existieren und sich deswegen viele Werte überlagern, wurde die Opazität der einzelnen Punkte verringert. Dadurch ist es weiterhin möglich zu erkennen an welchen Stellen sich Daten fokussieren. Da die Opazität aber nicht unendlich verringert werden kann ist weiterhin eine Funktion eingebaut welche beim hovern über einen Spieler anzeigt wie

viele andere Spieler von diesem überlagert werden. Die X und Y Achse sind immer mit dem aktuell ausgewählten Attribut beschriftet und skalieren bei Wechsel der Attribute automatisch mit. Abschließend sei zu erwähnen, dass eine Filterfunktion für Vereine und Nationalitäten der Spieler implementiert worden ist. Dadurch ist es den Anwendern möglich die Aufmerksamkeit auf spezifische Teilgruppen zu richten und mögliche lokale Trends in den Daten zu erkennen.

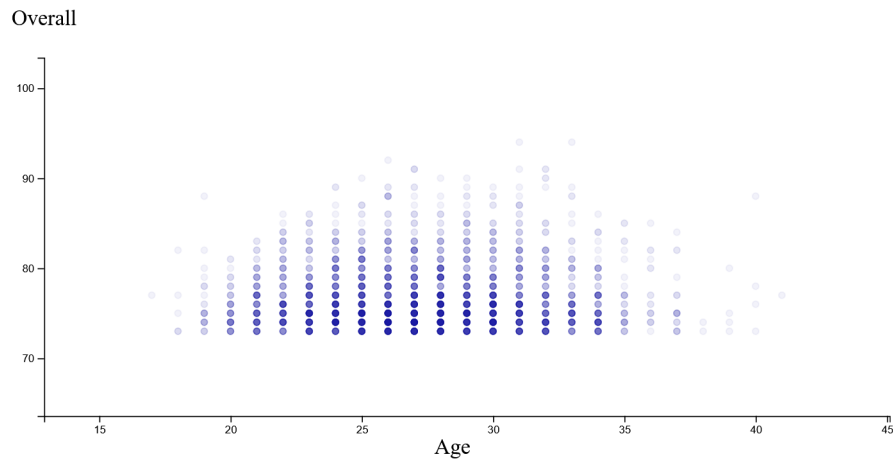


Abbildung 2: Darstellung des Scatterplots  
Quelle: eigene Darstellung

### 3.3.2 Visualisierung Zwei

Bei der zweiten Visualisierung handelt es sich um die der Parallelen Koordinaten. Wie bereits beschrieben ist diese Visualisierungsform dafür geeignet höherdimensionale Daten in einem zweidimensionalen Raum darzustellen. Zwar benötigt diese Darstellungsform mehr als einzelne Datensätze um interpretierbar zu sein, trotzdem muss ein einzelner Datensatz erkennbar bleiben. Aus diesem Grund sind zum einen Filterfunktionen für die Attribute: Verein, Nationalität und Position eingebaut worden, zum anderen ist eine *on: hover* Funktionalität implementiert worden, die den Pfad des ausgewählten Fußballers einfärbt, verbreitert und dessen Namen und Werte oberhalb der Darstellung anzeigt. Da es für die Parallelen Koordinaten relevant ist ob Attribute benachbart sind oder nicht, ist genauso wie bei der Scatterplot Darstellung die Option implementiert worden die Werte der Achsen vom Nutzer selbst per Knopfdruck festlegen zu lassen.

### 3.3.3 Visualisierung Drei

Bei der dritten Visualisierung handelt es sich um die explizite Baumdarstellung. Diese Darstellungsform ist besonders dazu geeignet Hierarchien zu visualisieren. Die Hierarchie hat in diesem Fall drei Ebenen: Die Fussballliga, die Vereine und die Spieler. Die Visualisierung hätte um Länder und Kontinente erweitert werden können, dies hätte aber keinen erheblichen Mehrwert



## 4 Implementierung

Die entwickelte Anwendung besteht aus drei verschiedenen Modulen, einem für jede Visualisierungstechnik. Dabei bauen alle drei Module auf den jeweiligen Übungen auf. Sie sind allerdings zum Teil stark erweitert und verändert worden. Besonders der Umstand, dass die Daten nicht direkt im Quellcode mit vorliegen sorgt dafür, dass die Vorlagen aus der Übung modifiziert werden müssen. Im Fall des Scatterplots hat das Model 3 Zustände: *Failure*, *Loading* und *Success*. Der *Success* Zustand des Models enthält dabei einen Record mit den geladenen Daten sowie zusätzlichen, für die Interaktion der Visualisierung, relevanten Variablen. Dieser grundlegende Aufbau sowie ein *Csv.Decoder* sind dabei aus einer anderen Übungen verwendet und angepasst worden. Besonders wichtig für die Funktionalität der Interaktionen sind die verschiedenen Zustände von *Msg*. Diese übergeben den vorherigen Zustand aller Attribute des bereits beschriebenen Records. Dabei wird lediglich der, der Funktion entsprechende Wert verändert. An der Darstellung des Scatterplots sind lediglich kleinere Veränderungen vorgenommen worden. Des Weiteren ist der Code um einige Ästhetische Funktionen sowie die Interaktionsmöglichkeiten erweitert worden.

Die bereits im Fall des Scatterplots beschriebenen Anpassungen des Übungscodes treffen ebenso auf die Parallelen Koordinaten zu. Weitergehend sind diese um eine zusätzliche Filterfunktion erweitert worden. Außerdem ist das einfärben des Pfades, sowie die Anzeige des Spielernamens sowie seiner Werte implementiert worden.

Der Code des Treediagramms ist im Vergleich zu dem der Übung am stärksten verändert worden. Da das Laden einer *.json* dazu führt, dass die Darstellung in der ELM Anwendung nicht mehr verändert werden kann, ist sich dazu entschieden worden die Daten wie in den anderen beiden Visualisierungen als *.csv* zu laden. Diese Daten können dann nach Bedarf gefiltert oder angepasst werden bevor sie vom *json.encoder* encoded werden. Ist dies geschehen kann wieder auf die Vorleistung der Übung zurückgegriffen werden und die Daten können durch den *TreeDecoder* decoded werden. Da in den ursprünglichen Daten keine Fussballligen vorliegen sind die verwendeten Ligen als Listen im ELM Code enthalten. Abschließend ist ein eigenes TreeLayout geschrieben worden, da bei Verwendung des vorgegebenen Treelayouts Informationen der Darstellung verloren gegangen wären.

**Wie ist der Quellcode gegliedert, was lies sich aus den Übungen übernehmen, Wie sieht die Datenstruktur des Modells aus -> in dem verschiedene Zustände der Interaktion gespeichert wurden (Success record)**

## 5 Anwendungsfälle

Im nachfolgenden Kapitel werden zu jeder der drei Visualisierungen ein konkreter Anwendungsfall vorgestellt. Zusätzlich werden Vor- und Nachteile der jeweiligen Visualisierungstechnik dis-

kutiert.

Spezifischen Anwendungsfall für Nutzergruppen vorstellen, der an Hand der Visualisierungstechniken visuell erkennbar ist, wäre dies auch mit anderen Techniken möglich gewesen? Aufwand mit anderen Techniken vergleichen

## 5.1 Anwendung Visualisierung Eins

Mit der Visualisierung des Scatterplots werden in Abbildung 5 die Attribute *Overall*<sup>2</sup> und *Wage* mit einander verglichen. Beispielhaft soll der Nutzen dieser Visualisierung anhand eines Spie-

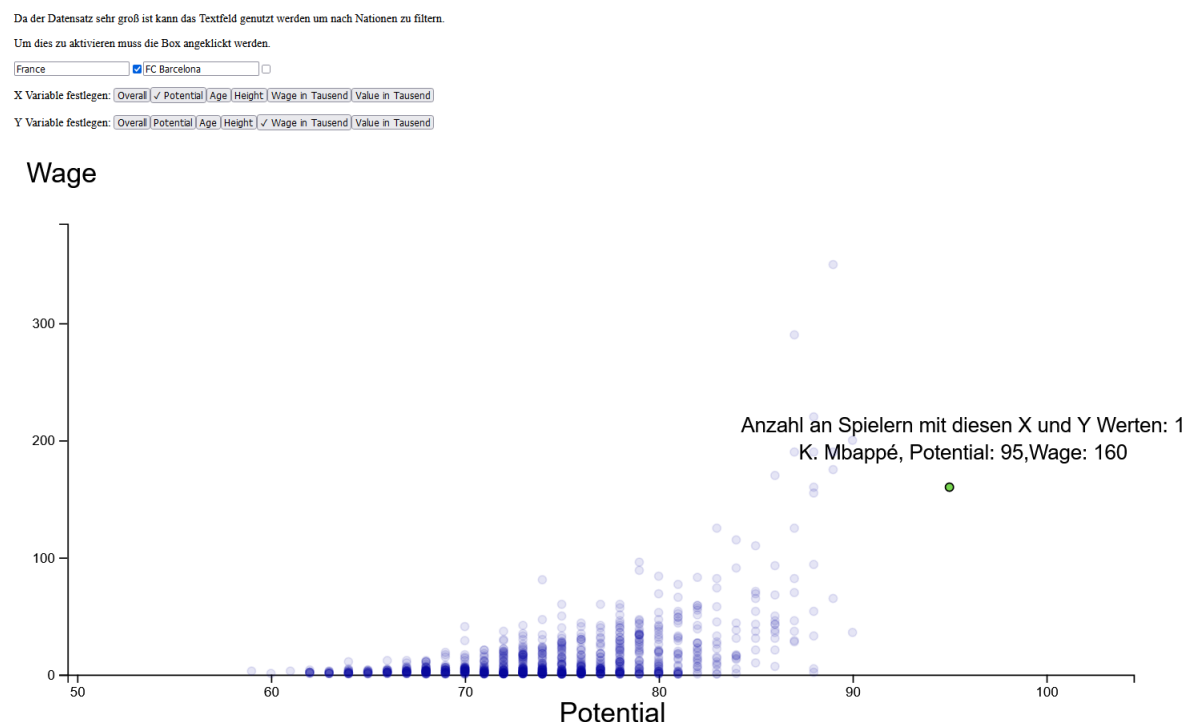


Abbildung 5: Darstellung des Scatterplot und dessen *on: hover* Funktionalität  
Quelle: eigene Darstellung

lers des Karrieremodus von FIFA 21 dargestellt werden. Dieser Nutzer möchte für sein Team im Karrieremodus einen neuen Spieler finden. Dieser soll ein höchstmögliches Potential haben. Trotzdem soll das Gehalt des Fussballers bezahlbar sein. Weiterhin soll der Spieler Franzose sein. Aus diesem Grund vergleicht der Nutzer die Attribute *Potential* und *Wage* mit dem zusätzlichen Nationalitätenfilter *France*. Dabei ist zu erkennen, dass mit steigendem Rating ab einem bestimmten Punkt das Gehalt exponential ansteigt. Ein Spieler, der zwar bereits ein sehr hohes Gehalt aufweist, für sein *Potential* von 95 aber trotzdem vergleichsweise günstig ist, ist Kylian Mbappé. Dieser Spieler ist ein Ausreißer aus dem generellen Trend der exponentiell steigenden *Wages* im Vergleich zum *Potential*.

<sup>2</sup>Bewertung eines Fussballers auf einer Skala von 1 bis 99

## 5.2 Anwendung Visualisierung Zwei

In der Visualisierung der Parallelen Koordinaten werden in Abbildungen 6 und 7 die Attribute *Defending*, *Physical*, *Pace* und *Passing* mit einander verglichen.

Da der Datensatz sehr groß ist können die Textfelder genutzt werden um nach Nationen, Clubs und Positionen zu filtern.

Um dies zu aktivieren muss die Box angeklickt werden.

Nationenfilter: Portugal ☐

Clubfilter: Juventus ☐

Positionsfilter: CB ☒

• Wert der ersten Spalte festlegen: Overall | Potential | Age | Pace | Shooting | Passing | Dribbling | Defending | Physical

• Wert der zweiten Spalte festlegen: Overall | Potential | Age | Pace | Shooting | Passing | Dribbling | Defending | Physical

• Wert der dritten Spalte festlegen: Overall | Potential | Age | Pace | Shooting | Passing | Dribbling | Defending | Physical

• Wert der vierten Spalte festlegen: Overall | Potential | Age | Pace | Shooting | Passing | Dribbling | Defending | Physical

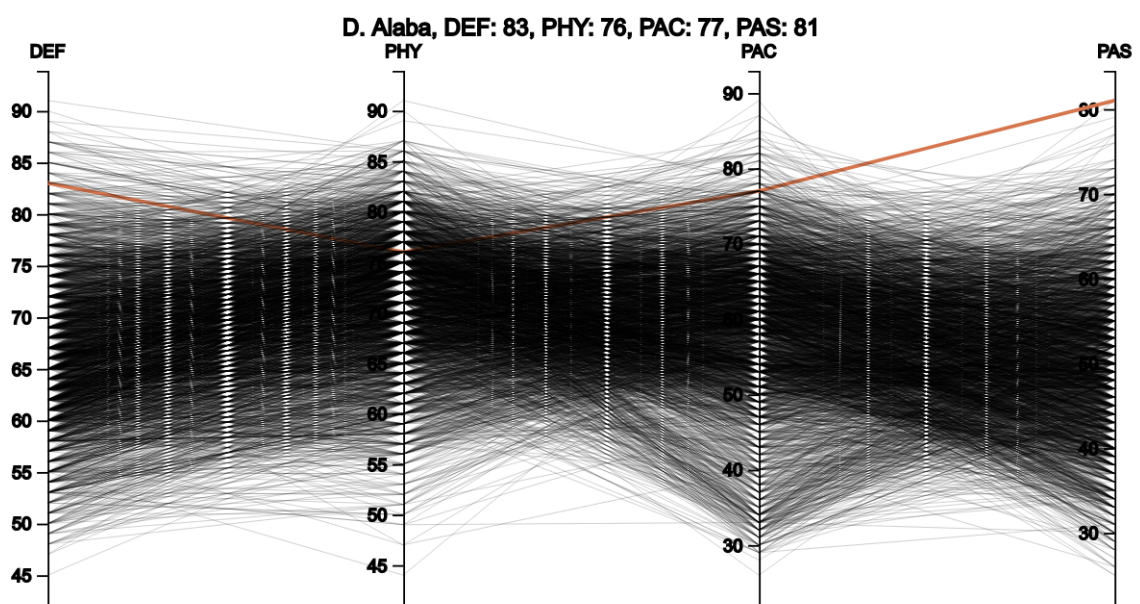


Abbildung 6: Darstellung der Parallelen Koordinaten inklusive *on:hover* Funktionalität  
Quelle: eigene Darstellung

Beispielhaft soll der Nutzen dieser Visualisierung anhand der Anforderungen eines Spielers des FUT Modus von FIFA 21 dargestellt werden. Dieser Anwender möchte sein Ultimate Team um einen Innenverteidiger (*Centerback: CB*) erweitern. Neben den für einen Innenverteidiger typischerweise wichtigen Attribute interessiert diesen Spieler auch die Passgenauigkeit des Innenverteidigers. Deswegen wählt er für den Vergleich die Attribute *Defending*, *Physical*, *Pace* und *Passing* aus. Generell ist bei Innenverteidigern in dieser Darstellung zu erkennen, dass die Passgenauigkeit im Vergleich zur Geschwindigkeit eher abfallend ist. Wird Abbildung 7 betrachtet in der Passen um zwei Positionen in der Darstellung nach Links verschoben worden ist, ist zu sehen, dass David Alaba entgegen des Trends - Passen ist im Vergleich zu Verteidigen und Physis eher geringer - einen höheren Passwert aufweist. Da die anderen Werte im Vergleich zu anderen Innenverteidigern trotzdem hoch liegt bietet sich die Auswahl dieses Innenverteidigers

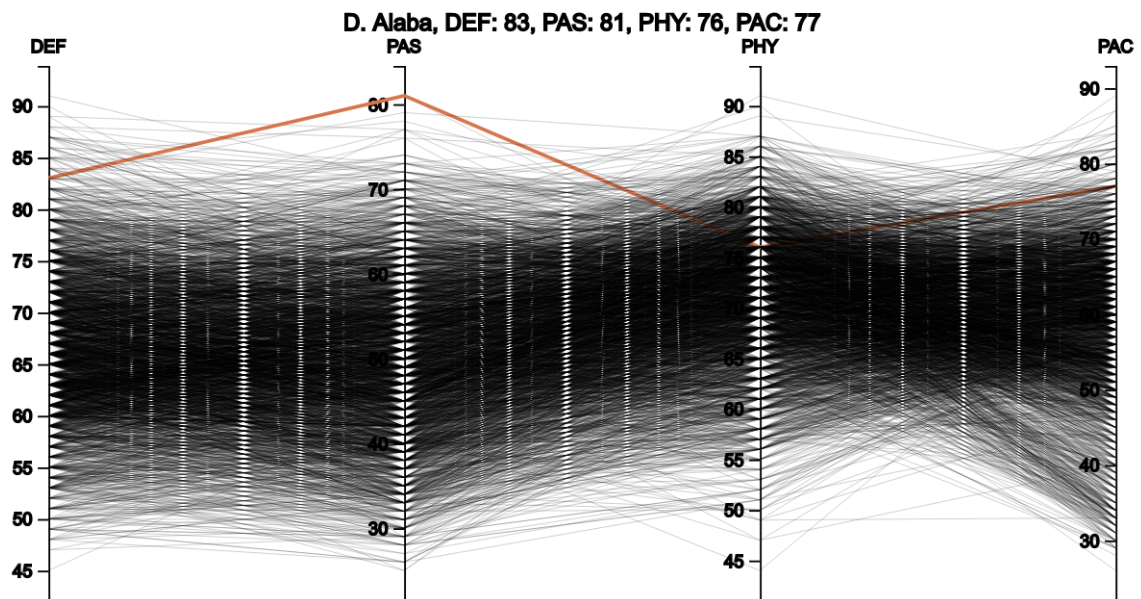


Abbildung 7: Darstellung der Parallelen Koordinaten inklusive *on: hover* Funktionalität  
Quelle: eigene Darstellung

für den Beispielnutzer an.

### 5.3 Anwendung Visualisierung Drei

In der Visualisierung der Baumdarstellung wird die Hierarchie von Bundesligavereinen inklusive Spielern der Position Rechtsverteidiger (Rightback: RB) gezeigt.

Beispielhaft wird diese wieder aus Sicht eines Spielers des FUT Modus betrachtet. In diesem Modus existiert ein Chemiesystem. Nach diesem System spielen nur Spieler gleicher Nation oder gleicher Liga gut miteinander. Weiterhin spielen Spieler gleicher Vereine besser zusammen. Ein Spieler ist dabei sich ein Team aus Spielern der Bundesliga zu bauen. Dabei fehlt allerdings ein Rechtsverteidiger. Um sich einen Überblick über die Rechtsverteidiger der Bundesliga zu verschaffen nutzt er die Funktionalität des Baumdiagramms so werden ihm alle vorhandenen Bundesliga Rechtstverteidiger angezeigt. Ist es ihm wichtig, dass der Spieler für einen spezifischen Verein spielt so kann auch dies mit Hilfe dieser Visualisierungstechnik erkannt werden.

## 6 Verwandte Arbeiten

Da es sich bei den verwendeten Daten um die eines Videospiels handelt existieren wenige wissenschaftliche Visualisierungen zu diesen. Allerdings existieren bereits einige nicht wissenschaftliche Visualisierungen, die hier kurz erwähnt werden sollen.

Zum einen ist dies die *FIFA Data Visualization*[5] einer Visualisierung des FIFA 19 Da-



Im Testfeld kann nach Positionen der Spieler gefiltert werden.  
 Um dies zu aktivieren muss die Box angeklickt werden.  
 ☒  
 Positionen zur Auswahl: GK, CB, LB, RB, CDM, CM, CAM, LW, RW, LM, RM, CF, ST  
 Wähle eine Liga aus:

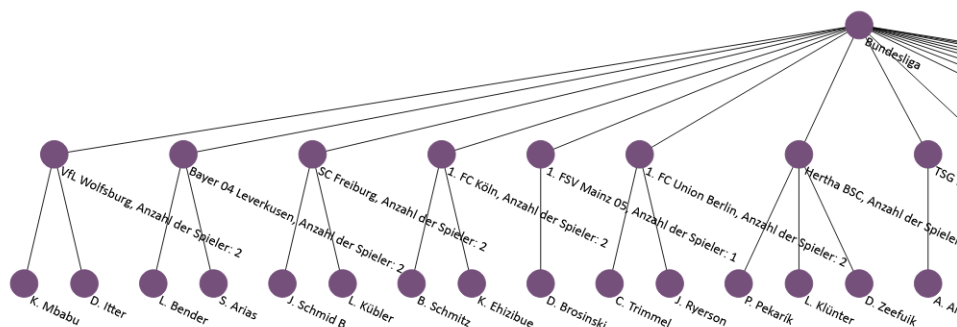


Abbildung 8: Darstellung des Baumdiagramms  
 Quelle: eigene Darstellung

tensatzes durch Roshan Sharma. In dieser Analyse der Daten werden verschiedene Werte des Datensatzes vor allem als Balken oder Kreisdiagramme dargestellt. Weiterhin werden Histogramme dargestellt. Im Verlauf der Analyse werden auch, in dieser Veranstaltung konkreter behandelte Darstellungsformen wie bspw. der Scatterplot oder die Darstellung als Sternförmige Koordinaten, verwendet.

Eine nichtwissenschaftliche Visualisierung des, in dieser Arbeit verwendeten, FIFA 21 Datensatzes ist von Paramartha Sengupta durchgeführt worden. In *FIFA 21: EDA and Visualization*[6] analysiert Sengupta vor allem unter Einsatz von Scatterplots und Balkendiagrammen. Weiterhin benutzt auch er die Sternförmigen Koordinaten um Attribute der Spieler zu vergleichen. Dabei beziehen sich die Analysen allerdings hauptsächlich auf aggregierte Daten und nicht auf Spieler spezifisch. Allerdings werden auch zwei Analysen mittels eines Scatterplots durchgeführt, in der ersten wird Alter von Spielern mit ihrem Overall verglichen in der zweiten wird das Alter der Spieler mit ihrem Potential verglichen. Beide Analysen sind auch mit dem hier vorgestellten Scatterplot möglich.

Um in der wissenschaftlichen Literatur vergleichbare Arbeiten zu finden muss sich vom Thema entfernt werden, zwar sind zum Videospiel FIFA keine wissenschaftlichen Artikel zu finden, jedoch lässt sich die hier durchgeführte Visualisierung mit der von Janetzko et. al.[3], beruhend auf realen Fussballdaten, vergleichen. In dieser Arbeit wird die Darstellungsform der Parallelen Koordinaten eingesetzt um Bewegungsdaten von Fussballern zu analysieren. Interessant ist dabei vor allem, dass die Darstellung aufgrund von Problemen der Erkennung von Clustern innerhalb der Achsen angepasst worden ist und um eine Farbkomponente erweitert worden ist. Weiterhin

sind die Interaktionsmöglichkeiten im Vergleich zu denen dieser Arbeit deutlich unterschiedlich. Es kann bspw. die Ober- und Untergrenze der Achsen verändert werden um so genauer in einzelne Cluster reinzoomen zu können. Weiterhin kann die Darstellung durch andere Visualisierungen überlagert werden. Zwar lässt sich sagen, dass die Visualisierung gegenüber der aus dieser Arbeit einige Vorteile bietet. Es muss jedoch erwähnt werden, dass der Fokus bezüglich Zielgruppen der Arbeiten und damit auch den Anforderungen unterschiedlich liegt. In der vorgestellten Arbeit liegt der Fokus vor allem auf der Erkennung von Clustern innerhalb der Gesamtmenge der Daten. Im Gegensatz dazu liegt der Fokus dieser Arbeit darauf Ausreißer nach oben aus den vorhandenen Clustern zu identifizieren um so die bestmöglichen Spieler zu finden. Trotzdem wäre die Erweiterung des Achsenskalierens möglicherweise bei einer Verbesserung der Visualisierungen in Betracht zu ziehen.

## **7 Zusammenfassung und Ausblick**

Die Visualisierungen haben es ermöglicht aus einem unübersichtlichen Datensatz mit über 16000 Zeilen Informationen herauszufiltern und anschaulich darzustellen. Anwendern ist es nach kurzer Einarbeitung möglich die Interaktionsmöglichkeiten der Visualisierungen selbst zu bedienen und auf neue Erkenntnisse zu stoßen. Dabei kann dies sowohl zum Sammeln von Informationen über einzelne Spieler, als auch zur Analyse größerer Trends getan werden. Besonders die Darstellungsform der Parallelen Koordinaten bietet im Vergleich zur Analyse der Daten als Liste einen echten Mehrwert, da Anwendern nicht nur die Werte der einzelnen Spieler angezeigt werden sondern durch die Pfade aller Spieler ein Maßstab für die Einschätzung der Werte einzelner Fussballer geschaffen wird.

## Anhang: Git-Historie

Das die Commits von zwei verschiedenen Accounts stammen ist zu entschuldigen. Die Anmeldung in der GIT GUI von Windows ist unabsichtlich mit einem anderen Account vorgenommen worden und erst später bemerkt worden.

- \* ba194f5 (HEAD -> master, origin/master) (Writing Chapter 2, 2021-09-09)
- \* a32805e (Changes to html testsite, 2021-09-09)
- \* fcfa8fa (Updates to Scatterplot, 2021-09-08)
- \* 9f41152 (Added additional filterfunctionality to parallelCoordinates, 2021-09-08)
- \* 130d43d (Updated Readme.md, 2021-09-08)
- \* bcb2233 (changes to spData.csv reagrding encoding, 2021-09-08)
- \* 14007ce (Added filter functionality to parallel Coordinates, 2021-09-08)
- \* b373216 (deleted unused datasets, 2021-09-08)
- \* 75f726f (started writing chapter preprocessing, 2021-09-08)
- \* a2207bf (Renaming of elm files and Creation of elm files with smaller csv files, 2021-09-08)
- \* a2e9801 (fixed error in index.html, 2021-09-08)
- \* da0991c (Included new html file for parallel COordinates, 2021-09-08)
- \* 9f20f40 (Updated RDataPreprocessing file, 2021-09-08)
- \* c72d4a2 (Updated spData and spData2000 with value and wage, 2021-09-08)
- \* e46740a (test changes in spData.csv, 2021-09-08)
- \* 956bf26 (test update spDate, 2021-09-08)
- \* 379c0de (updated spData again, 2021-09-08)
- \* 477c43a (Updated spData.csv with wage and value data, 2021-09-08)
- \* 3c200e1 (Updated Baumdata.csv with Premier League Data, 2021-09-07)
- \* 3af31ee (Updated Baum csv data with serie a players, 2021-09-07)
- \* 1e66fcd (Development of Filterfunction for Players Positions in Treediagram, 2021-09-07)
- \* 1cfd18f (Updated Baumdata.csv, 2021-09-07)
- \* f5adf5f (reworking of 1.1, 2021-09-07)
- \* 1bd9293 (Creation of Tree.html, 2021-09-07)
- \* b04c990 (Created index.html for github pages, 2021-09-07)
- \* ebfba7f (Updated bibtex file, 2021-09-07)
- \* c6240a4 (First try of preprocessing parallel coordinates data, 2021-09-06)
- \* 43ba73b (Added highlighting and Player info on hover, 2021-09-06)
- \* 447a93e (changed additional errors in baumData.csv reagrding laLiga players, 2021-09-06)
- \* 30992b9 (fixed errors in LaLiga club names, 2021-09-06)
- \* 3a8659a (Updated Baumdata.csv with laliga clubs, 2021-09-06)
- \* 7a98e59 (Started writing preprocessing file in R, 2021-09-06)
- \* d4b3df2 (Changed error in Baumdata regarding multiple club names, 2021-09-06)
- \* c825fef (Created test csv Data for Baumdiagram, 2021-09-06)

- \* 79150e1 (Implementation of Filter function in the Scatterplot as well as changes to the
- \* d3bc365 (-Created Fifa 21 Scatterplot Data, reduced to 2000 Players, 2021-09-05)
- \* 605e204 (Added preliminary Scatterplot Data from Fifa 21, 2021-09-05)
- \* cce366d (-Reorganization and inclusion of placeholder images, 2021-09-05)
- \* a6328bf (Added full csv of fifa 21 data to gitignore, 2021-09-05)
- \* abbce42 (started writing chapter 2.2, 2021-09-05)
- \* 1bcf321 (-Reorganized git 2 deleting additional, unused files, 2021-09-05)
- \* 0017063 (-Reorganized git, 2021-09-05)
- \* e5ebc82 (Updated Readme.me, 2021-09-05)
- \* 93b8a97 (-Started writing Chapter 2.1, 2021-09-05)
- \* 4477dba (Started writing chapter 1.3, 2021-09-05)
- \* 7be9390 (Started writing Chapter Zielgruppen, 2021-09-04)
- \* 2386040 (Reduced size of csv for testing purposes, 2021-08-22)
- \* 5d98177 (Added Nationality to CSV, 2021-08-22)
- \* a395085 (Added smaller CSV tree file, 2021-08-15)
- \* 4fa41ce (New Test CSV File for ElmTreeMap, 2021-08-15)
- \* 5952879 (Added .json file and Excercise tree visualization code to start implementing t
- \* 1295900 (-Changed name of tex and pdf file and added new folder AbgabeOrdnerLatex, 2021
- \* 5c1747d (-Changed gitignore to new Latex files, 2021-08-13)
- \* 9443014 (Kapitel 2 Daten Version 0.1, 2021-08-12)
- \* f673cd6 (-Grundlagen des Projektberichts, 2021-08-12)
- \* cd852b2 (TXT statt csv, 2021-08-12)
- \* 9c07724 (-Added smaller CSV with 1000 lines, 2021-08-12)
- \* a441b52 (-csv, 2021-08-12)
- \* 2f7fbe0 (-csv, 2021-08-12)
- \* 8ac06b8 (-minor fixes csv, 2021-08-12)
- \* cc9cede (Prefiltering of Data, 2021-08-12)
- \* eb68618 (Testing error with height, 2021-08-12)
- \* 95935c0 (Testing Height in Metres instead of CM, 2021-08-12)
- \* ef3a227 (Rounded Height, 2021-08-12)
- \* cb4c956 (Reduced CSV file, 2021-08-12)
- \* f0e6cae (Added Interactivity to parallel Coordinates, 2021-08-11)
- \* 9ee66e1 (Added Parallele Koordinaten Grundfunktionalität, 2021-08-11)
- \* eb4f173 (-Added Git Folder structure, 2021-08-11)
- \* 7d09347 (Added changeablity to the data, 2021-08-10)
- \* 07422d2 (Extended number of encodeable variables for the Scatterplot, 2021-08-06)
- \* a6fd816 (-Added possibility to change variables, 2021-08-03)
- \* f0f68ea (-Updated PointName -Started writing Paper, 2021-08-03)
- \* adfd7a6 (-First working Scatterplot, 2021-08-03)

```

* c4f4b5f (-Scatterplot first tries, 2021-08-03)
* f9af25c (-Committing new files, 2021-07-30)
* 4618601 (Upload pdf, 2021-07-07)
* d267f3d (-deleted Latex files, 2021-07-07)
* 65fd4fd (Merge branch 'master' of https://github.com/Blauekuh/Visualisierung_FIFA19.g
|\
| * 68629b2 (Update .gitignore, 2021-07-07)
* | 7aaf7f3 (Latex File created, 2021-07-07)
|/
* be4234d (-Added gitignore, 2021-07-05)
* 546ae0e (Upload CSV Dataset for Fifa 19, 2021-06-17)
* c640421 (Commit of first test elm file, 2021-06-16)
* fcdc3ad (origin/main, origin/Development) (Initial commit, 2021-06-16)

```

## Literatur

- [1] *A Complete Guide to Scatter Plots*. en-US. URL: <https://chartio.com/learn/charts/what-is-a-scatter-plot/> (besucht am 09.09.2021).
- [2] Stephen Few. “Multivariate Analysis Using Parallel Coordinates”. en. In: (), S. 9.
- [3] Halldór Janetzko u. a. “Enhancing Parallel Coordinates: Statistical Visualizations for Analyzing Soccer Data”. In: *Visualization and Data Analysis 2016, San Francisco, California, USA, February 14-18, 2016*. Hrsg. von David L. Kao, Thomas Wischgoll und Song Zhang. Ingenta, 2016, S. 1–8. URL: <http://ist.publisher.ingentaconnect.com/contentone/ist/ei/2016/00002016/00000001/art00011>.
- [4] *Min-Maxing*. de-DE. Feb. 2014. URL: <https://mein-mmo.de/lexikon/min-maxing/> (besucht am 04.09.2021).
- [5] Sharma Roshan. *FIFA Data Visualization*. en. URL: <https://kaggle.com/roshansharma/fifa-data-visualization> (besucht am 27.08.2021).
- [6] Paramartha Sengupta. *FIFA 21: EDA and Visualization*. en. URL: <https://kaggle.com/paramarthasengupta/fifa-21-eda-and-visualization> (besucht am 28.08.2021).