

MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG

Projektbericht zum Modul Information Retrieval und Visualisierung im

Sommersemester 2021

Designstudie zur Visualisierung von Daten des Videospiels FIFA 21

Eingereicht bei: Dr. Alexander Hinneburg

Eingereicht von: Johannes Lange

Git-Repository: https://github.com/JohannesLange/Visualisierung_FIFA19

Eingereicht am: 12. September 2021

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Abkürzungsverzeichnis	III
1 Einleitung	1
1.1 Anwendungshintergrund	2
1.2 Zielgruppen	2
1.3 Überblick und Beiträge	4
2 Daten	4
2.1 Technische Bereitstellung der Daten	5
2.2 Datenvorverarbeitung	5
3 Visualisierungen	6
3.1 Analyse der Anwendungsaufgaben	6
3.2 Anforderungen an die Visualisierungen	7
3.3 Präsentation der Visualisierungen	7
3.3.1 Visualisierung Eins	7
3.3.2 Visualisierung Zwei	8
3.3.3 Visualisierung Drei	9
3.4 Interaktion	9
4 Implementierung	10
5 Anwendungsfälle	10
5.1 Anwendung Visualisierung Eins	10
5.2 Anwendung Visualisierung Zwei	11
5.3 Anwendung Visualisierung Drei	13
6 Verwandte Arbeiten	14
7 Zusammenfassung und Ausblick	15
Anhang: Git-Historie	IV

Abbildungsverzeichnis

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-cf	3
2	Darstellung des Scatterplots und dessen <i>on:hover</i> Funktionalität Quelle: eigene Darstellung	11
3	Darstellung der Parallelen Koordinaten inklusive <i>on:hover</i> Funktionalität Quelle: eigene Darstellung	12
4	Abbildung 3 mit nach Links verschobenem <i>Passing</i> Quelle: eigene Darstellung . .	12
5	Darstellung des Baumdiagramms Quelle: eigene Darstellung	13

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 - stetig an Bedeutung zu. Um aus großen Mengen von Daten neue Informationen gewinnen zu können, reicht es oft nicht aus, die Daten mit rein mathematischen Methoden der Statistik zu analysieren. Durch den Einsatz 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 sind dazu häufig Guides verwendet worden, in denen hunderte Seiten Tipps und Tricks abgedruckt worden sind. Diese Guides bieten jedoch weder Interaktion noch können sie jeden Datensatz eines Spiels beschreiben. Zusätzlich sind sie oft aus der subjektiven Sicht eines einzelnen Autors verfasst.

Moderne Visualisierungstechniken erlauben es, deutlich mehr Informationen zu extrahieren und für die Spieler interaktiv nutzbar zu machen. Aus diesem Grund sollen in dieser Designstudie drei Visualisierungstechniken vorgestellt werden, die sich dafür eignen, den gegebenen Datensatz zur Erkenntnisgewinnung zu visualisieren.

Der vorliegende 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, Attribute wie *Name*, *Alter*, *Verein*, *Nationalität*, *Größe* und *Gewicht* 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. Dadurch besteht die Hoffnung mittels Visualisierungstechniken noch weitgreifendere Erkenntnisse aus dem gegebenen Datensatz ziehen zu können, als solche die lediglich Fans des Spiels dienen.

Mögliche Anforderungen an die Visualisierungen von Spielern des Spiels könnten dabei bspw. die Folgenden sein: Welcher Fussballer verbessert das Team langfristig, kostet aber vergleichsweise wenig. Bsp.: *Welche Spieler hat das höchste Potential bei geringstem Wage?*

Eine Weitere Frage könnte sich beim Bau eines Teams bestehend aus Bundesliga Vereinen ergeben, wenn ein Spieler einer bestimmten Position benötigt wird. Bsp.: *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?*

Offen bleibt, ob Fragen von anderen Anforderungsgruppen, wie bspw. Sportwissenschaftlern, beantwortet werden können: *Wann erreichen Fussballer ihren Peak? Sind größere Fussballer besser, schneller, etc.?* Da alle Daten, die nicht konkret der Realität entnommen werden können, von Mitarbeitern des Herstellers EA Sports festgelegt worden sind, können diese nicht ohne weitere Analyse genutzt werden. 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 können Trends in Daten aufgezeigt werden, wie z.B. ob sich das Alter von Fussballern auf deren Bewertung im Spiel auswirkt. Weiterhin können so Spieler gefunden werden, deren Werte von den gefundenen Trends abweichen.

Die zweite Visualisierungstechnik ist die der Parallelen Koordinaten.[2] Diese eignet sich aufgrund ihrer Beschaffenheit dazu, mehrdimensionale Daten in einem zweidimensionalen Raum auf übergreifende Trends zu untersuchen. Trotzdem kann sie auch genutzt werden, um spezifische Fussballer konkret zu untersuchen und deren Werte im Vergleich zu den beschriebenen Trends zu betrachten. Bspw. kann der Frage nach der Eignung eines Fussballers zum Einsatz auf einer spezifischen Position nachgegangen werden. 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, bei automatischer Achsenskalierung, nur eine gerade Linie an. Weiterhin lassen sich lediglich Zusammenhänge zwischen benachbarten Attributen erkennen. Dies kann bei ungeschulten Anwendern jedoch schnell zu Fehlinterpretationen führen.

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 einfachsten zu verstehen.

1.2 Zielgruppen

Da es sich bei den Daten um die eines Videospiels handelt, ist davon auszugehen, dass eine der Zielgruppen 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 sechs verschiedene Werte der Fussballer von Relevanz (siehe Werte in Abbildung 1). Dadurch, dass *FUT* zu den Onlinespielen gehört, in denen viele Spieler sich durch *Min-Maxing*[6] einen Vorteil verschaffen wollen, ist bei den Anwendern von einem hohen Vorwissen bezüglich der Daten auszugehen. Mit Blick auf die Visualisierungstechniken ist bei dieser Zielgruppe nicht davon auszugehen, dass die Funktionsweise der Parallelen Koordinaten bekannt ist. Trotzdem wird dieser Gruppe genug Interesse unterstellt, um bereit zu sein, Zeit in das Verständnis einer Visualisierungstechnik zu investieren. Aus diesem Grund bietet sich diese Technik trotzdem an. Eine weitere - für diese Gruppe relevante Visualisierungstechnik - ist die explizite

Baumdarstellung. Durch diese ist es möglich, zu untersuchen, welche Fussballer in welcher Liga und für welchen Verein spielen. Dadurch können Spieler dabei unterstützt werden, ihr Team passend zusammenzustellen. Die Funktionsweise eines Baumdiagramms sollte dabei den meisten Spielern bereits bekannt 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 *Wage*, *Age*, *Overall* oder *Potential* ist dabei relevant. In diesem Fall bietet sich eine Darstellung per Scatterplot an. Mit dieser Technik lassen sich schnell die aussichtsreichsten Fussballer erkennen. 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 haben und deswegen 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 für diese Gruppe die Visualisierung mittels eines Baumdiagramms an. Dadurch kann sich ein grober Überblick über Spieler einer Liga gemacht werden. Durch die Möglichkeit des Filterns nach Positionen können Spieler auch grob 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 von Scatterplots sowie Baumdiagrammen sollte den meisten Mitgliedern dieser Zielgruppe allerdings bereits bekannt sein.



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-stcf>

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 und parallele Koordinaten könnten nach Analyse der Daten auf Realitätsnähe dazu genutzt werden mögliche Fragestellungen zu beantworten. Weiterhin ist zu erwähnen, 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, werden Guides zu Videospielen bisher meist in Textform veröffentlicht 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. Dadurch ist es möglich den gesteigerten Informationsbedarf zu decken ohne direkt Programmierkenntnisse beim Spieler vorauszusetzen.

Für die Visualisierungstechnik Scatterplot ist sich vor allem aufgrund dessen Verbreitung und Simplizität entschieden worden. Da es sich bei den möglichen Zielgruppen zum Großteil nicht um Wissenschaftler handelt, bietet der Scatterplot eine verständliche Visualisierung an, die aber trotzdem einen deutlichen Mehrwert gegenüber den Daten in rein tabellarischer Form aufweist. Das Baumdiagramm ist aufgrund der Möglichkeit hierarchische Daten darstellen zu können, ausgewählt worden. Durch dieses kann übersichtlich gezeigt werden, welche Spieler in welchen Vereinen und welche Vereine in welcher Liga spielen.

Im Gegensatz zu diesen beiden Visualisierungen sind die Parallelen Koordinaten unbekannter und von Laien nicht direkt zu verstehen. Trotzdem gehören sie zu den verständlichsten Techniken, um multidimensionale Daten in einem zweidimensionalen Raum darzustellen. Durch sie wird vor allem der interaktive Vergleich von mehr als zwei Attributen ermöglicht. Wie genau die Visualisierungen umgesetzt worden sind und welchen Mehrwert jede einzelne Technik gegenüber anderen Visualisierungsformen bietet, wird in Kapitel 3 diskutiert. Wie Anforderungen spezifischer Nutzer erfüllt werden, ist in Kapitel 5 beschrieben.

2 Daten

Bei den Daten handelt es sich um einen Datensatz des Fussball Videospiels Fifa 21¹. Er ist der Webseite Kaggle entnommen und enthält Daten zu 107 Attributen von über 16000 Fussballern des Videospiels FIFA 21. In diesen enthalten sind - neben Attributen wie Name, Verein und Alter - auch eine ungefähr nach der Qualität des Fussballers festgelegte Bewertung sowie spezifische Informationen zu den fussballerischen Fähigkeiten (Schießen, Passen, Dribbling, Verteidigen, Geschwindigkeit, Physis, etc.) der Spieler.

Grundsätzlich eignen sich die Daten gut, um die genannten Fragestellungen beantworten zu können. Allerdings enthält der Grunddatensatz einige Felder, die für die Visualisierung nicht nötig sind. Aus diesem Grund ist der Datensatz in der Vorvorarbeitung verkleinert worden (siehe 2.2). Da der Datensatz nicht nur im Sinne der Attribute pro Fussballer sehr groß ist, sondern auch bezüglich der Anzahl an Fussballern werden die verwendeten Visualisierungen schnell unübersichtlich. Um dies zu verhindern, ist entschieden worden, zusätzliche Filterfunktionen einzubauen.

¹<https://www.kaggle.com/ekrembayar/fifa-21-complete-player-dataset> besucht am 10.09.2021

2.1 Technische Bereitstellung der Daten

Der Originaldatensatz ist aufgrund seiner Größe und der fehlenden Möglichkeit, einige der numerischen Werte als solche auslesen zu können, nicht direkt verwendet worden (siehe Kapitel 2.2). Aus diesem Grund wird er nicht mit bereitgestellt, ist aber unter folgendem Link zu finden: <https://www.kaggle.com/ekrembayar/fifa-21-complete-player-dataset>

Die verwendeten Daten sind in einem öffentlichen Github Repository gehostet². Dort liegen die Daten der drei Visualisierungen jeweils als *.csv* im *data* Ordner vor. Die Trennung der Daten zwischen den Visualisierungen ist getätigt worden, 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.

2.2 Datenvorverarbeitung

Nicht alle Attribute sind für die Visualisierung relevant. Deswegen, und weil ein kleinerer Datensatz die Visualisierung beschleunigt, sind nur die Attribute im verwendeten Datensatz enthalten, die Relevanz für die Visualisierung haben. Da sich die relevanten Attribute auch zwischen den verschiedenen Visualisierungstechniken unterscheiden, sind für jede Anwendung unterschiedliche Datensätze vorverarbeitet und gefiltert worden.

Die Vorverarbeitung der Daten ist in *R* durchgeführt worden. Das verwendete Skript (*RData-Preprocessing.r*) liegt im Github vor. Im ersten Schritt der Vorverarbeitung 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 numerischen Wert 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 komplexer, da die Zahlen mit *M* für Millionen und *K* für Tausend abgekürzt sind. Diese Werte sind *in Tausend* im finalen Datensatz enthalten. Werte, die im Datensatz mit einem *M* versehen sind, werden mit 1000 multipliziert; 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). Nachdem diese Werte angepasst sind, können die für Scatterplot und Parallele Koordinaten relevanten Werte in jeweils neuen *Data.frames* gespeichert und als *.csv* wieder exportiert werden.

Für die Darstellung des Baumdiagramms ist sich auf die vier größten europäischen Fussballligen (Premier League, Bundesliga, La Liga und Serie A) konzentriert worden. Da im Datensatz keine Informationen zur Liga vorliegen, sind die relevanten Spieler über eine manuell erstellte Liste der Vereine dieser Ligen gefiltert worden. Aus diesem Grund enthält der Datensatz dieser

²https://github.com/JohannesLange/Visualisierung_FIFA19/tree/master

Visualisierungstechnik nur Daten von Spielern, deren Verein Teil einer dieser vier Ligen ist. Im Git Repository sind im Ordner *data* die vorverarbeiteten Datensätze (Scatterplot: *spData.csv*, Parallele Koordinaten: *pkData.csv*, Baumdiagramm: *baumData.csv*) 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 Bedürfnissen der Zielgruppen ergeben, eingegangen. Weiterhin werden im Verlauf des Kapitels 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 verborgen bleiben würden, oder nur mit erheblichem Zeitaufwand zu erkennen wären. Dabei bieten Scatterplot und Parallele Koordinaten die Möglichkeit, größere Trends innerhalb der Daten zu erkennen: *Korrelieren Alter und Bewertung eines Fussballers? Sind teurere Fussballer besser?* Aber auch Detailanforderungen wie die Suche nach einem Innenverteidiger, der die Wünsche des Nutzers bestmöglich erfüllt, lassen sich mit Hilfe der Parallelen Koordinaten befriedigen: *Welcher Innenverteidiger hat einen besonders hohen Passwert bei akzeptablen Bewertungen der anderen relevanten Attribute?*

Durch die eingebaute Funktion, dass der Pfad eines Spielers *on: hover* 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.

Ähnliches bewirkt die Funktion der Farbänderung und Namensanzeige *on: hover* über Punkten des Scatterplots. Dazu ist zu erwähnen, dass diese Funktion auch bei der Analyse größerer Trends hilfreich sein kann. Da sich viele diskrete Werte des Datensatzes überschneiden, ist die Opazität der Punkte zwar bereits verringert worden. Eine weitere Reduzierung würde einzelne Punkte allerdings fast unsichtbar machen. Trotzdem ist durch die Größe des Datensatzes die maximale Opazität an an einigen Stellen schnell erreicht. Aus diesem Grund wird *on: hover* neben dem Namen des Spielers auch die Anzahl an Spielern, die von diesem überlagert werden, angezeigt. Im Vergleich zu diesen Techniken erfüllt das explizite Baumdiagramm andere Anforderungen. Mit ihm lassen sich Hierarchien darstellen. In Kombination mit der implementierten Filterfunktion lassen sich so bspw. Fragen nach Spielern bestimmter Positionen in den Ligen beantworten. Insgesamt ist zu beachten, dass die Zielgruppen mit den Visualisierungsformen Scatterplot und Baumdiagramm vermutlich bereits in Kontakt gekommen sind, und diese deswegen bereits interpretieren können. Parallele Koordinaten hingegen sind eine vergleichsweise unbekannte Visualisierungstechnik. Aus diesem Grund bietet es sich an, den Parallelen Koordinaten eine Erklärung

der Funktionsweise beizulegen. Dadurch kann den Nutzern geholfen werden diese Technik verstehen und nutzen zu können.

3.2 Anforderungen an die Visualisierungen

Wie bereits im vorherigen Abschnitt beschrieben worden ist, adressieren die Visualisierungen spezifische Ziele. Um diesen Zielforderungen gerecht zu werden, müssen sie auch die sich ergebenden Anforderungen erfüllen. In diesem Abschnitt werden deswegen die Designentscheidungen, welche sich aus den Anforderungen ergeben, erklärt.

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 es, 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 erfüllt 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 im Fall der Parallelen Koordinaten eine *on: hover* Funktion implementiert worden, die den gewählten Punkt oder Pfad hervorhebt, und Name sowie Informationen zu dem jeweiligen Fussballer anzeigt.

Um den Nutzern einen Mehrwert gegenüber analogen Visualisierungen zu bieten und den Bedarf selbstgewählter Vergleiche von Daten zu decken, ist in die Visualisierungen des Scatterplots und der Parallelen Koordinaten die Funktion eingebaut worden, die zu vergleichenden Attribute selbst zu wählen.

Da Daten in Baumdiagrammen bei einer großen Menge an Kindern schnell unübersichtlich werden können, ist im Baumdiagramm eine Filterfunktion nach Ligen und Positionen der Spieler implementiert worden. So kann sich fast ohne Scrollen ein Überblick über alle Spieler einer Position verschafft werden.

3.3 Präsentation der Visualisierungen

Im folgenden Kapitel werden die drei Visualisierungsformen vorgestellt. Weiterhin wird auf Funktionalität und Interaktionsmöglichkeiten genauer eingegangen. Zusätzlich werden Vor- und Nachteile der jeweiligen Visualisierungstechnik gegenüber möglichen Alternativen diskutiert.

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 enthalten sind, und deswegen viele Überlagerungen existieren,

ist die Opazität der einzelnen Punkte verringert worden. Dadurch ist es weiterhin möglich, zu erkennen, an welchen Stellen sich Daten fokussieren. Da die Opazität aber nicht unendlich verringert werden kann, ist zusätzlich eine Funktion eingebaut, die beim *hovern* mit der Maus über einem Punkt anzeigt, wie viele andere Fussballer von diesem überlagert werden. Da diese Funktion die Laufzeit der Visualisierung deutlich erhöht, ist unter dem Namen *ScatterplotFast.elm* eine Version des Codes ohne diese Funktion enthalten. Die X und Y Achse sind immer mit den vom Nutzer ausgewählten Attributen beschriftet, und skalieren bei Wechsel der Attribute automatisch mit.

Zusätzlich ist 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.

Betrachtet man den Scatterplot im Vergleich zu anderen zweidimensionalen Visualisierungstechniken - wie beispielsweise dem Balkendiagramm - zeigt sich ein großer Vorteil des Scatterplots: Obwohl übergreifende Trends in den Daten zu erkennen sind, lassen sich die Einzelobservationen noch voneinander differenzieren. Dieser Fakt ist in einigen der beschriebenen Anwendungsfälle definitiv relevant.

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* Funktion implementiert worden, die den Pfad des ausgewählten Fussballers einfärbt, verbreitert und dessen Namen und Werte oberhalb der Darstellung anzeigt. Für die Nutzung der Parallelen Koordinaten ist es relevant ob Attribute benachbart sind oder nicht. Aus diesem Grund ist, genauso wie bei der Scatterplot Darstellung, die Option implementiert worden, die Werte der parallelen Achsen vom Nutzer selbst per Knopfdruck festlegen zu lassen. Eine ähnliche Visualisierungsform zur zweidimensionalen Darstellung mehrdimensionaler Daten ist das Sterndiagramm. Zwar bietet dieses Vorteile - wie beispielsweise den Fakt, dass kein Attribut am Ende allein ohne Nachbarn steht. Außerdem werden die Achsen meist manuell skaliert, so haben einzelne Observationen Aussagekraft, ohne einen Vergleich mit anderen zu benötigen. Trotzdem ist die Verwendung der Parallelen Koordinaten bei den gegebenen Anforderungen die beste Wahl. Bei einem Datensatz mit der Größe des hier gegebenen wären die Daten im Sterndiagramm nicht zu erkennen. Generell ist diese Darstellungsform eher geeignet, um wenige Daten zu vergleichen und nicht große Datenmengen auf Trends zu untersuchen.

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 ist allerdings nicht getan worden, da den Zielgruppen dadurch kein Mehrwert entstanden wäre.

Explizite Bäume sind eine Visualisierungstechnik, die schnell in die Breite wachsen und unübersichtlich werden können. Deswegen ist neben der Funktion einzelne Ligen auswählen zu können noch eine Filterfunktion nach Positionen implementiert worden. So wird die Darstellung entweder aller Spieler einer Liga oder aller Spieler einer bestimmten Position einer Liga ermöglicht.

Ist das Visualisierungsziel das Darstellen von Hierarchien, so gibt es verschiedene Techniken zur Auswahl. Diese reichen von Sunburstdiagrammen über implizite Baumdiagramme bis hin zu den zur Zeit sehr beliebten Sankeydiagrammen. Trotzdem ist das explizite Baumdiagramm noch immer die leichtverständlichste und übersichtlichste Visualisierung hierarchischer Daten. Zwar wird ein Baumdiagramm bei einer großen Menge an Kinddaten schnell sehr breit, trotzdem sind die Daten zu erkennen. Sunburstdiagramme, Sankeydiagramme und implizite Baumdiagramme bieten zwar alle den Vorteil, dass die Größe ihrer Elemente zusätzliche Informationen übermitteln, haben kleine Elternelemente allerdings viele Kinder, sind diese oft kaum noch zu erkennen. Dieser Fall kann bei expliziten Baumdiagrammen nicht eintreten. Da durch die Anforderungen lediglich die hierarchische Darstellung, nicht aber Anzeigen bestimmter Mengenverhältnisse relevant ist, ist die explizite Baumdarstellung gewählt worden.

3.4 Interaktion

Die Visualisierungen enthalten die im vorherigen Kapitel bereits beschriebenen Interaktionsmöglichkeiten. Diese sind im Fall von Scatterplot und Parallelen Koordinaten bspw. Buttons, durch welche die Attribute der Achsen ausgewählt werden können. Zusätzlich existieren Texteingabefelder mit denen die Daten gefiltert werden können. Eine weitere Interaktion ist die *on: hover* Funktionalität, durch die es Nutzern möglich ist, spezifische Spieler aus der Menge an Daten herauszusuchen. Das Baumdiagramm hat vergleichsweise weniger Interaktionsmöglichkeiten. Es ist ebenso möglich, Daten über eine Texteingabe zu filtern und bestimmte Daten per Button auszuwählen. Weitere Interaktionen existieren allerdings nicht.

Die Darstellungen sind technisch voneinander unabhängig. Es existieren keine Interaktionen, die visualisierungsübergreifend funktionieren. Da die Anwendungszwecke der Techniken sich jedoch sowieso stark voneinander unterscheiden, würde eine solche Interaktion vergleichsweise wenig Mehrwert bieten und die einzelnen Visualisierungen möglicherweise unverständlicher machen.

4 Implementierung

Die entwickelte Anwendung besteht aus vier verschiedenen Modulen, einem für jede Visualisierungstechnik und eine schnellere Version des Scatterplots. Dabei bauen alle vier Module auf den Grundlagen der 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.

Das *Model* aller vier Module hat deswegen 3 Zustände: *Failure*, *Loading* und *Success*. Der *Success*-Zustand des Models enthält dabei einen *Record*. Dieser *Record* enthält die geladenen Daten sowie zusätzliche - für die Interaktion der Visualisierung relevante - 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. Lediglich der zu ändernde Wert wird durch die Funktion angepasst. So können Änderungen am Modell vorgenommen werden, ohne die Daten neu laden zu müssen. An der Optik des Scatterplots sowie der Parallelen Koordinaten sind lediglich kleinere Veränderungen vorgenommen worden. Allerdings ist der Code um die bereits beschriebenen Interaktionsmöglichkeiten erweitert worden.

Der Code des Baumdiagramms 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 *Tree.Decoder* 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 gehen würden.

5 Anwendungsfälle

Im nachfolgenden Kapitel wird zu jeder der drei Visualisierungen ein konkreter Anwendungsfall vorgestellt. Dadurch wird der Mehrwert, den die jeweilige Visualisierung bietet, aufgezeigt.

5.1 Anwendung Visualisierung Eins

In der Visualisierung des Scatterplots werden in Abbildung 2 die Attribute *Overall*³ und *Wage* mit einander verglichen. Beispielhaft soll der Nutzen dieser Visualisierung anhand eines Spielers

³Bewertung eines Fussballers auf einer Skala von 1 bis 99

des Karrieremodus von FIFA 21 dargestellt werden.

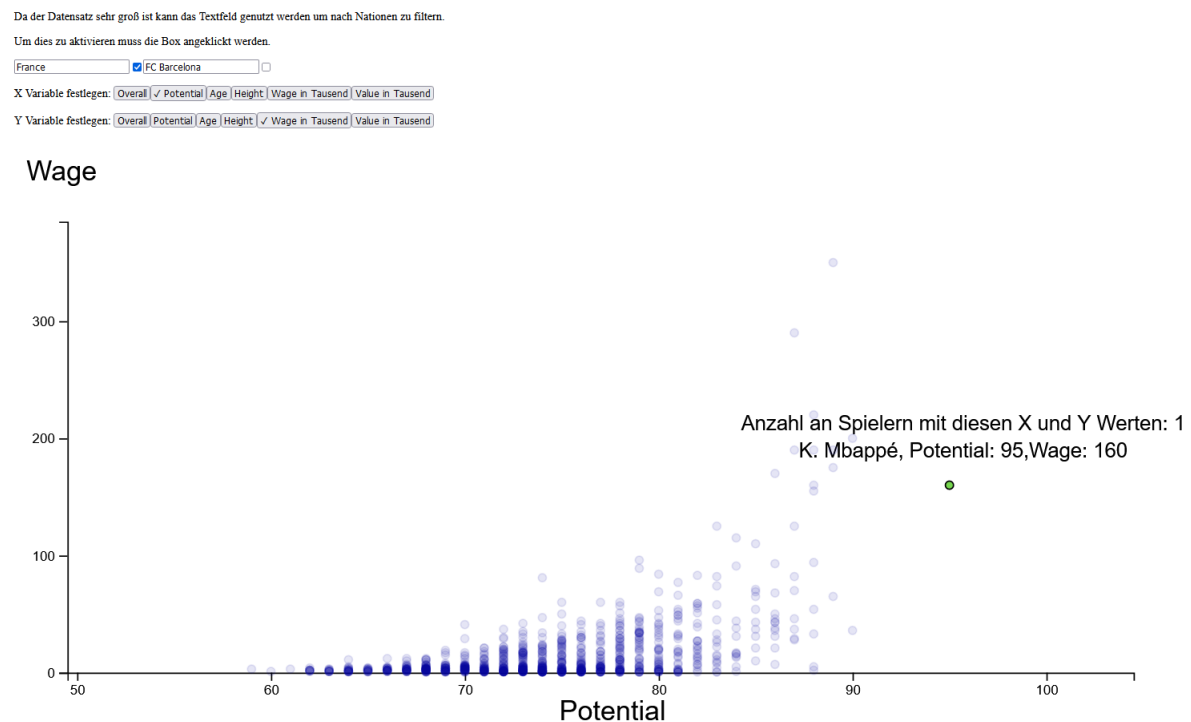


Abbildung 2: Darstellung des Scatterplots und dessen *on: hover* Funktionalität
Quelle: eigene Darstellung

Dieser Nutzer möchte für sein Team im Karrieremodus einen neuen Spieler finden, der ein höchstmögliches Potential haben soll. 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 *Potential* ab einem bestimmten Punkt, *Wage* exponentiell 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*. Eine günstigere Option wäre in diesem Fall beispielsweise Dayot Upamecano, dessen Potential liegt zwar nur bei 90, das Gehalt allerdings auch nur bei 36000€. Dem Nutzer der Anwendung stehen so mehrere Optionen offen. Das Abweichen dieser Fussballer von Trends in den Daten wird erst durch diese Visualisierung offensichtlich.

5.2 Anwendung Visualisierung Zwei

In der Visualisierung der Parallelen Koordinaten werden in Abbildungen 3 und 4 die Attribute *Defending*, *Physical*, *Pace* und *Passing* miteinander 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: ☐

Clubfilter: ☐

Positionsfilter: ☒

• Wert der ersten Spalte festlegen:

• Wert der zweiten Spalte festlegen:

• Wert der dritten Spalte festlegen:

• Wert der vierten Spalte festlegen:

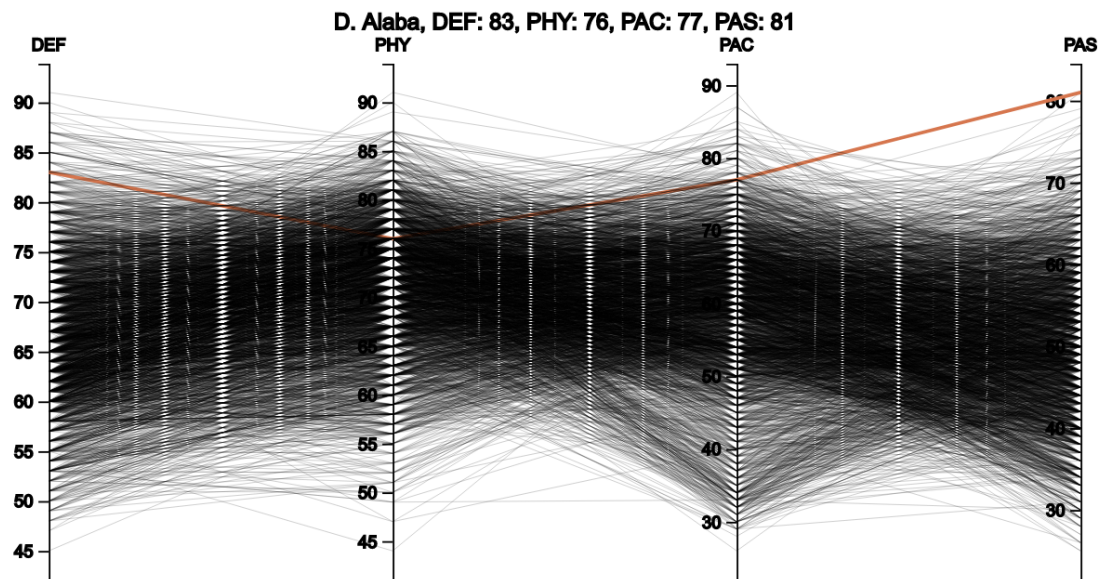


Abbildung 3: Darstellung der Parallelen Koordinaten inklusive *on: hover* Funktionalität
Quelle: eigene Darstellung

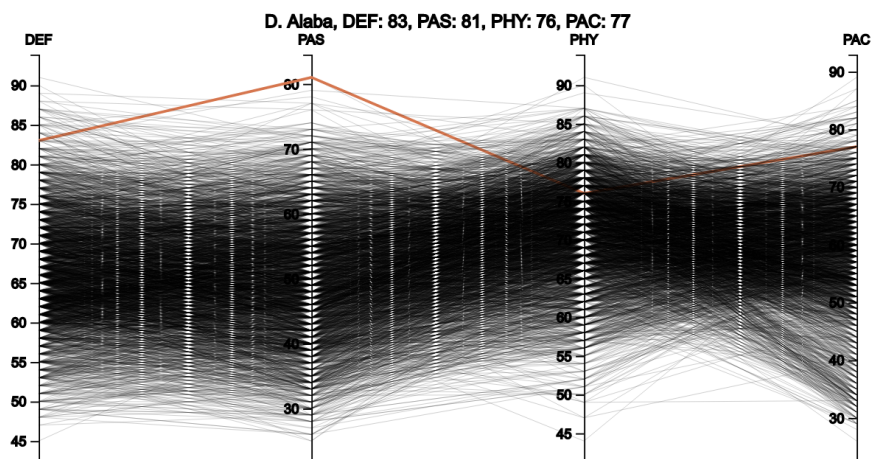


Abbildung 4: Abbildung 3 mit nach Links verschobenem *Passing*
Quelle: eigene Darstellung

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 relevanten Attributen 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 4 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 trotz dessen hoch liegen, bietet sich die Auswahl dieses Innenverteidigers 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.

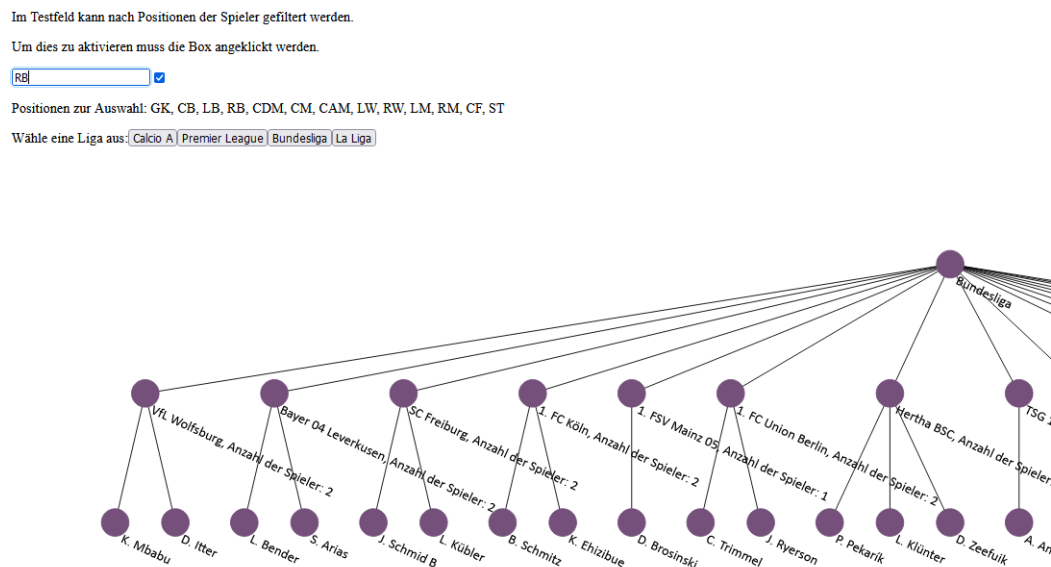


Abbildung 5: Darstellung des Baumdiagramms
Quelle: eigene Darstellung

Beispielhaft wird diese wieder aus Sicht eines Spielers des FUT Modus betrachtet. In diesem Modus existiert ein Chemiesystem.[3] Nach diesem System spielen nur Spieler gleicher Nation oder gleicher Liga gut miteinander. Weiterhin spielen Fussballer gleicher Vereine besser zusammen.

Ein Nutzer ist dabei, sich ein Team aus Spielern der Bundesliga zu erstellen. Dabei fehlt allerdings ein Rechtsverteidiger. Um sich einen Überblick über die Rechtsverteidiger der Bundesliga zu verschaffen, nutzt er die Funktionalität des Baumdiagramms. Dadurch werden ihm alle vorhandenen Bundesliga Rechtsverteidiger angezeigt. Ist es ihm wichtig, dass der Spieler für einen spezifischen Verein spielt, so kann auch dies mit Hilfe dieser Visualisierungstechnik erkannt werden. Da die Menge der Kindelemente in dieser Anforderung keine Rolle spielt, bietet sich Verwendung der expliziten Baumdarstellung im Vergleich zu den anderen in Kapitel 3 diskutierten Techniken zur Darstellung hierarchischer Strukturen an.

6 Verwandte Arbeiten

Da es sich bei den verwendeten Daten um die eines Videospiele 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*[7] eine Visualisierung des FIFA 19 Datensatzes durch Roshan Sharma. In dieser Analyse der Daten werden verschiedene Werte des Datensatzes überwiegend als Balken oder Kreisdiagramme dargestellt. Weiterhin werden Histogramme verwendet. Im Verlauf der Analyse werden auch in dieser Veranstaltung konkreter behandelte Darstellungsformen - wie bspw. der Scatterplot oder die Darstellung als Sternförmige Koordinaten - verwendet. Dabei liegt der Fokus auf der Darstellung von Mengenverteilungen verschiedener Attribute.

Eine nichtwissenschaftliche Visualisierung des in dieser Arbeit verwendeten FIFA 21 Datensatzes ist von Paramartha Sengupta durchgeführt worden. In *FIFA 21: EDA and Visualization*[8] analysiert Sengupta den Datensatz vor allem unter Einsatz von Scatterplots und Balkendiagrammen. Weiterhin werden Sternförmige Koordinaten verwendet, um Attribute der Spieler zu vergleichen. Dabei beziehen sich die Analysen allerdings hauptsächlich auf aggregierte Daten und nicht auf spieterspezifische. Allerdings werden mittels eines Scatterplots zwei Analysen durchgeführt, die so auch mit dem Scatterplot in dieser Arbeit durchgeführt werden können. In der ersten wird *Age* von Spielern mit ihrem *Overall* verglichen. In der zweiten wird das *Age* der Spieler mit ihrem *Potential* verglichen.

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.[4] - beruhend auf realen Fussballdaten - vergleichen. In dieser Arbeit wird die Darstellungsform der Parallelen Koordinaten eingesetzt, um Bewegungsdaten von Fussballern zu analysieren. Aufgrund von Problemen bei der Erkennung von Clustern innerhalb der Achsen, ist die Darstellung um eine Farbkomponente erweitert worden. Weiterhin unterscheiden sich die Interaktionsmöglichkeiten im Vergleich zu denen dieser Arbeit deutlich. 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 auch 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. Bei einer möglichen Erweiterung dieser Arbeit wäre die Möglichkeit des manuellen Achsenskalierens trotzdem in Betracht zu ziehen.

Um die in der Einleitung gestellte Frage: *Können die gegebenen Daten und Visualisierungen genutzt werden um auf reale Sportler rückschließen zu können?* beantworten zu können werden die Visualisierungen mit den Daten von Kalén et. al.[5] verglichen. In der Diskussion des Papers wird beschrieben, dass der Peak der Spielerleistung ungefähr im Alter von 31 Jahren erreicht wird. Dieses Erkenntnis wird beim Betrachten des Scatterplots (X-Achse: *Age*, Y-Achse: *Overall*) ebenso erreicht, da in der Darstellung mit steigendem *Age* ein steigendes *Overall* zu erkennen ist. Dieses stagniert dann um das Alter von 31 und fällt danach wieder leicht ab. Somit lässt sich die Vermutung anstellen, dass die Daten auch Sportwissenschaftlern behilflich sein können. Um dies bestätigen zu können, müsste allerdings eine, den Umfang dieser Arbeit überschreitende, Analyse der Daten durchgeführt werden. Aus diesem Grund sind die Daten mit Blick auf Rückschlüsse zu realen Sportlern mit Vorsicht zu betrachten. Lügen Realdaten vor, würden sich die Visualisierungen hingegen definitiv zur deren Darstellung eignen.

7 Zusammenfassung und Ausblick

Die Visualisierungen haben es ermöglicht, aus einem unübersichtlichen Datensatz mit über 16000 Zeilen Informationen herauszufiltern, und anschaulich darzustellen. Anwenden 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 Anwenden 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. Zusammenfassend sind vor allem die Möglichkeit die Daten selbst filtern zu können, selbst Attribute der Achsen festzulegen sowie die *on: hover* Funktionalität die großen Vorteile, die von den Visualisierungen geboten werden.

Mit Blick auf mögliche Verbesserungen der Visualisierungen fällt beim Scatterplot zum einen ein, statt die Menge an gleichen Daten durch Opazität von einander zu unterscheiden, dies mit

Hilfe einer Farbskala zu tun. Ein weiterer Punkt, der verbessert werden kann, ist der Fakt, dass bei Überlagerung der Daten immer nur der Name eines Spielers angezeigt wird. Das Anzeigen einer Liste der Fussballer, die die gleichen Werte haben, wäre in diesem Fall eine Verbesserungsmöglichkeit.

Die Parallelen Koordinaten könnten, wie bereits in Kapitel 6 erwähnt, um die Möglichkeit des manuellen Skalierens der Achsen erweitert werden, um so das Hereinzoomen in Cluster zu ermöglichen. Im Fall des Baumdiagramms wäre es denkbar, dieses um eine Funktion zu erweitern, in der mehr Attribute der Spieler präsentiert werden. So könnten Anwender mehr Nutzen aus dieser Visualisierung ziehen. Das Hinzufügen zusätzlicher Fussballligen wäre eine weitere sinnvolle Erweiterung. Abschließend ist zu sagen, dass trotz der Verbesserungsmöglichkeiten bereits durch die vorhandenen Visualisierungen ein echter Mehrwert gegenüber den Rohdaten entsteht.

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.

- * a0a626f (HEAD -> master, origin/master) (Finished third rewrite, 2021-09-12)
- * 64b220c (fixed typo in Link, 2021-09-12)
- * f518954 (Finished second rewrite, 2021-09-11)
- * 1445f0b (Updated typo, 2021-09-11)
- * 96e6511 (Updated the html files for
the github.io Website, 2021-09-11)
- * 07907a4 (Updated Readme.md, 2021-09-11)
- * d96e885 (started second rewrite, 2021-09-11)
- * 320e368 (deletion of unused .csv files, 2021-09-11)
- * c131c04 (Restructuring of the git repository, 2021-09-11)
- * 0cb2bfe (Minor changes to .tex, 2021-09-11)
- * 441f403 (First rewrite Chapter 7, 2021-09-11)
- * a725be3 (Small changes to elm code (reorganizing)
Additional faster Scatterplot without aggregation function, 2021-09-11)
- * 0dd7525 (First rewrite Chapter 4-6, 2021-09-11)
- * f094598 (First rewrite Chapter 3.3.3, 2021-09-11)
- * b988614 (first rewrite - Chapter 3.3.1, 2021-09-10)
- * 9376d8d (rewriting of Chapter 1, 2021-09-10)
- * e657a51 (Started writing Chapter 6 and 7, 2021-09-10)
- * 7658876 (Started writing Chapter 5, 2021-09-10)
- * 11a1ed0 (Started writing Chapter 4, 2021-09-09)
- * d0d1119 (Upload of test data to improve Scatterplot speed, 2021-09-09)
- * febf2dd (Reworking of chapters 1 and 3, 2021-09-09)
- * 735a9a8 (Started writing Chapter 3, 2021-09-09)
- * c5b8f90 (rewriting of Chapter 1.2, 2021-09-09)
- * f4057eb (Attempt to change layer order in parallel Coordinates, 2021-09-09)
- * 8bb04b7 (Corrected Linking errors in .html files, 2021-09-09)
- * 8856893 (rewriting of Chapter 1.1 and
inclusion of additional bibtex entries, 2021-09-09)
- * d5546c0 (Updated Readme.me with additional info, 2021-09-09)
- * 903f6c8 (rewriting of Einleitung, 2021-09-09)
- * ba194f5 (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 C0ordinates, 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 selections, 2021-09-06)
- * 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 tree visualization, 2021-08-14)
- * 1295900 (-Changed name of tex and pdf file and added new folder AbgabeOrdnerLatex, 2021-08-13)
- * 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 (-Commiting 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.git, 2021-07-07)
- | \
- | * 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: *perceptual edge* (Sep. 2006). URL: http://www.perceptualedge.com/articles/b-eye/parallel_coordinates.pdf (besucht am 09.09.2021).
- [3] Niklas Heib. *FIFA Ultimate Team: So funktioniert die Team- und Spielerchemie in FUT*. de-DE. Juli 2021. URL: https://www.tonight.de/games/fifa/fifa-ultimate-team-fut-wie-funktioniert-chemie-teamchemie-spielerchemie-links-loyalitaet_77330.html (besucht am 07.09.2021).
- [4] 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>.
- [5] Anton Kalén u. a. "Are Soccer Players Older Now Than Before? Aging Trends and Market Value in the Last Three Decades of the UEFA Champions League". In: *Frontiers in Psychology* 10 (2019), S. 76. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2019.00076. URL: <https://www.frontiersin.org/article/10.3389/fpsyg.2019.00076> (besucht am 07.09.2021).
- [6] *Min-Maxing*. de-DE. Feb. 2014. URL: <https://mein-mmo.de/lexikon/min-maxing/> (besucht am 04.09.2021).
- [7] Sharma Roshan. *FIFA Data Visualization*. en. URL: <https://kaggle.com/roshansharma/fifa-data-visualization> (besucht am 27.08.2021).

- [8] Paramartha Sengupta. *FIFA 21: EDA and Visualization*. en. URL: <https://kaggle.com/paramarthasengupta/fifa-21-eda-and-visualization> (besucht am 28.08.2021).