

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

Eingereicht am: 9. September 2021

# Inhaltsverzeichnis

<b>Abkürzungsverzeichnis</b>	<b>III</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Anwendungshintergrund . . . . .	1
1.2 Zielgruppen . . . . .	2
1.3 Überblick und Beiträge . . . . .	3
<b>2 Daten</b>	<b>4</b>
2.1 Technische Bereitstellung der Daten . . . . .	4
2.2 Datenvorverarbeitung . . . . .	4
<b>3 Visualisierungen</b>	<b>5</b>
3.1 Analyse der Anwendungsaufgaben . . . . .	6
3.2 Anforderungen an die Visualisierungen . . . . .	6
3.3 Präsentation der Visualisierungen . . . . .	6
3.3.1 Visualisierung Eins . . . . .	6
3.3.2 Visualisierung Zwei . . . . .	7
3.3.3 Visualisierung Drei . . . . .	7
3.4 Interaktion . . . . .	7
<b>4 Implementierung</b>	<b>7</b>
<b>5 Anwendungsfälle</b>	<b>8</b>
5.1 Anwendung Visualisierung Eins . . . . .	8
5.2 Anwendung Visualisierung Zwei . . . . .	8
5.3 Anwendung Visualisierung Drei . . . . .	8
<b>6 Verwandte Arbeiten</b>	<b>8</b>
<b>7 Zusammenfassung und Ausblick</b>	<b>8</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> . . . . .	2
2	Darstellung des Scatterplots Quelle: eigene Darstellung . . . . .	6
3	Darstellung der Parallelen Koordinaten Quelle: Eigene Darstellung . . . . .	7
4	Ausschnitt aus Baumdarstellung Quelle: Eigene Darstellung . . . . .	7

## Abkürzungsverzeichnis

**FUT** Fifa Ultimate Team

# 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. In diesem Bericht geht es um die Visualisierung von Daten aus dem Computerspiel Fifa 21. Da sich diese Daten allerdings auf tatsächliche Fussballer beziehen besteht die Hoffnung mittels Visualisierung dieser Daten Rückschlüsse auf die Sportler selbst schließen zu können. So könnten Erkenntnisse aus den Daten deutlich wertvoller sein als zuerst vermutet.

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. Da nicht alle Eigenschaften eines Fussballers konkret festgehalten werden können, sind einige der Werte geschätzt. Da aus diesem Grund wie oben beschrieben nicht sicher auf die tatsächlichen Spieler zurückgeschlossen werden kann, bietet sich eine Analyse der Daten bezüglich der Nähe zur Realität an. Doch selbst wenn die Daten nicht zu neuen Erkenntnissen bezüglich der realen Sportler führen können sich für Spieler des Videospiels neue Erkenntnisse ergeben.

Konkrete Fragen für Spieler des Spiels könnten dabei sein: *Welche Spieler hat das höchste Overall bei geringstem Gehalt?* oder *Welche Linksverteidiger spielen in der Bundesliga?*

Weiterhin eignet sich gerade die Visualisierung mittels Scatterplot um Korrelation zwischen Daten zu untersuchen.

## 1.1 Anwendungshintergrund

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

Die zweite Visualisierungstechnik ist die der parallelen Koordinaten. Diese eignet sich auf Grund ihrer Beschaffenheit dazu mehrdimensionale Daten zu untersuchen. In diesem Fall wurde sich für parallele Koordinaten mit vier verschiedenen Attributen entschieden um eine gewisse Übersichtlichkeit zu wahren. 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. Dabei kann ein Überblick über Fussballer einer spezifischen Fussballliga gewonnen werden. Da diese Daten sehr schnell unübersichtlich werden kann sind die Daten nach Ligen gefiltert. Weiterhin können diese Daten noch nach Positionen gefiltert werden. So kann ein Spieler des FUT Modus, der für sein Team einen Linksverteidiger der Bundesliga benötigt nach dieser Liga und Position

filtern.

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 würde ich weiterhin in die des Spielmodus *Karriere* und die des Spielmodus *Fifa Ultimate Team FUT* unterteilen.

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 des 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.

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 sich diese nicht durch Training oder ähnliches ändern lassen. Weil dieser Modus sehr kompetitiv ist kann ein Fussballer bereits durch einen schlechten Wert in einer Kategorie für Spieler unbrauchbar werden. Dadurch, dass *FUT* zu den Onlinespielen gehört in denen viele Spieler sich durch *Min-Maxing*[2] einen Vorteil verschaffen wollen ist von einem hohen Vorwissen der Anwender der Visualisierung auszugehen. Für diese Spieler bieten sich vor allem die Visualisierung mittels der parallelen Koordinaten zum Vergleich mehrerer Variablen sowie die Darstellung mittels des Baumdiagramms an.



**Fifa 19 Spieler -> Viel Vorwissen,  
Fussball Fan -> recht viel Vorwissen,  
Sportinteressent -> Kaum Vorwissen**

Zielgruppen für Scatterplot:

Parallele Koordinaten:

Abbildung 1: Darstellung einer FUT Karte inklusive relevanter Werte am

Baumdiagramm: Hier lässt sich erkennen welche Sportliga Europas im Durchschnitt die besten Spieler hat. Deswegen ist dies für X interessant.

Mögliche Zielgruppen: Videospieler, Teilnehmer einer Fantasy Fußball Liga

### 1.3 Überblick und Beiträge

Bei den Daten handelt es sich um einen Datensatz des Fußball Videospiels Fifa 21 dieser enthält Daten zu über 16000 Fussballern.

In diesen enthalten sind neben Variablen wie Name, Verein und Alter auch ein ungefähr nach der Qualität des Fussballers festgelegtes Rating sowie Bewertungen der fussballerischen Fähigkeiten (Schießen, Passen, Dribbling, Verteidigen, Geschwindigkeit, Physis, etc.)

Die erste verwendete Visualisierungstechnik ist der Scatterplot, dieser ist gut dazu geeignet zwei Variablen der Fussballer zu vergleichen. Bspw. könnte untersucht werden ob ein Fussballer entsprechend seines Könnens (Rating) verdient (Wage). Für Spieler des Karrieremodus können gerade solche Informationen relevant sein, da so Spieler gefunden werden können, die besser spielen als ihr Gehalt es vermuten lassen würde oder solche die zu teuer für ihr Können sind.

Bei der zweiten verwendeten Visualisierungstechnik handelt es sich um Parallele Koordinaten. Diese, im Vergleich zum Scatterplot, etwas fortgeschrittene Technik eignet sich um mehr als zwei Werte miteinander zu vergleichen. Ihr Einsatz bietet sich daher an um Werte eines Fussballers in den Attributen: Schießen, Passen, Dribbling, Verteidigen, Geschwindigkeit oder Physis miteinander zu vergleichen und dies vorher nach der Position des Fussballers zu Filtern um zu untersuchen ob sich ein Spieler anhand seiner Attribute dafür eignet auf dieser Position zu spielen. (Bsp. ein Stürmer benötigt vor allem Geschwindigkeit, Schießen und Dribbling, ein Verteidiger hingegen Verteidigen, Physis und Geschwindigkeit). Diese Technik könnte besonders für Spieler des FUT Modus relevant sein, da sich so die besten Spieler für den kompetitiven Online-Modus finden lassen.

Die dritte Visualisierungstechnik ist die Darstellung als Baumdiagramm. durch sie werden Spieler ihren Teams und diese ihren jeweiligen Ligen zugeordnet. Da es im FUT Modus ein Chemie System gibt indem ein Team nur gut zusammenspielen kann wenn Spieler einer Liga oder eines Teams entstammen kann ein Nutzer der Visualisierung mit Hilfe dieser ein Team aus einer spezifischen Liga zusammen bauen und nach jeder einzelnen Position filtern[1].

Als dritte und letzte Technik habe ich die Baumdarstellung ausgewählt um sichtbar zu ma-

chen welche Fussballliga Europas im Durchschnitt die besten Spieler hat. Deswegen ist dies für X interessant.

## 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 den gleichen Werte aufweisen.

### 2.1 Technische Bereitstellung der Daten

Die Daten sind in einem privaten Github 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.

Auch die Daten der Visualisierung als Baumdiagramm liegen als *.csv* vor und werden erst im Programm zu einer *.json* encoded und dann wieder zu einem Baumdiagramm decoded. So liegen die Daten nicht starr als *.json* vor und können innerhalb des Programms vorgefiltert werden. 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.

---

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

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.

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

### 3 Visualisierungen

Der gegebene Datensatz wird auf drei verschiedene Arten visualisiert. Die erste Visualisierung ist ein Scatterplot. Mit diesem können zwei verschiedene Attribute des Datensatzes vom Nutzer freigewählt werden und die Spieler anhand dieser miteinander verglichen werden.

Bei der zweiten Visualisierungsform handelt es sich um parallele Koordinaten, diese erlauben im Vergleich zum Scatterplot den Vergleich von mehr als zwei Attributen, sind aber vergleichsweise unübersichtlicher. Weiterhin müssen Nutzer deutlich bewandelter sein um sie interpretieren zu können, da Zusammenhänge nicht benachbarter Attribute kaum erkenntlich sind und vom Nutzer erwartet wird dieser nach Rechts und Links zu verschieben.

Die dritte Visualisierung ist ein explizites Baumdiagramm. Mit diesem kann die Struktur der Fussballligen Europas dargestellt werden. Dabei können Nutzer die Liga auswählen und die Spieler nach ihren Positionen filtern.

Da der Datensatz sehr groß ist können die Spieler nach Verein und Nationalität noch zusätzlich gefiltert werden. Da es trotz dessen noch häufig Überschneidungen der Daten gibt, bspw. bei Alter und Overall der Spieler existiert eine Funktion, die anzeigt wie viele Spieler genau diese



Werte aufweisen. Weiterhin sind die angezeigten Punkte teilweise durchsichtig. Dadurch können Überschneidungen der Punkte auch durch deren Opazität erkannt werden.

### 3.1 Analyse der Anwendungsaufgaben

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

### 3.2 Anforderungen an die Visualisierungen

Wie muss die Visualisierung designed werden um das Zielproblem gut beantworten zu können?

### 3.3 Präsentation der Visualisierungen

Visualisierungstechniken vorstellen, Interaktivität zeigen, Designentscheidungen begründen(Erfüllen diese die Anforderungen?), Diskutieren wieso nicht andere Techniken verwendet wurden(Expressivität und Effektivität).

#### 3.3.1 Visualisierung Eins

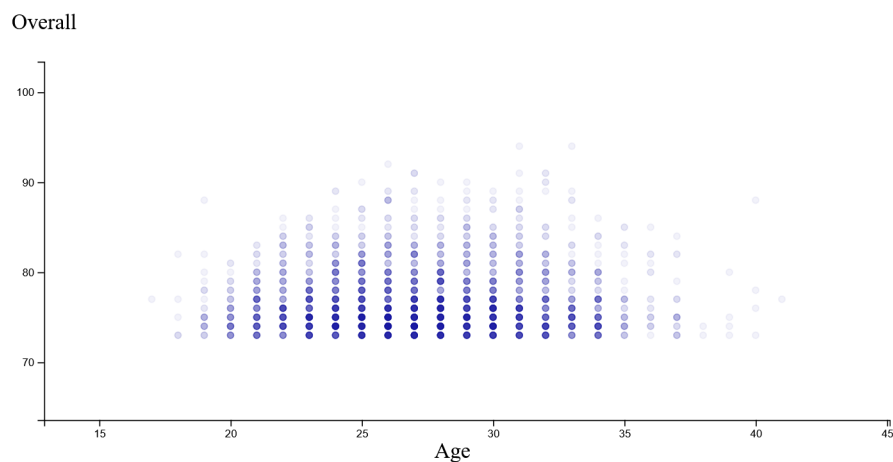


Abbildung 2: Darstellung des Scatterplots  
Quelle: eigene Darstellung

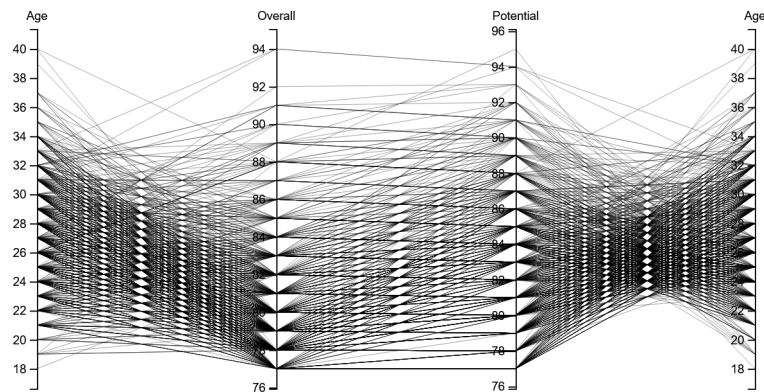


Abbildung 3: Darstellung der Parallelen Koordinaten  
Quelle: Eigene Darstellung

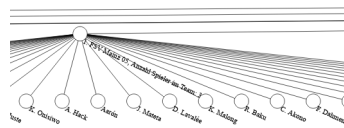


Abbildung 4: Ausschnitt aus Baumdarstellung  
Quelle: Eigene Darstellung

### 3.3.2 Visualisierung Zwei

### 3.3.3 Visualisierung Drei

### 3.4 Interaktion

Interaktionen in den Visualisierungen(möglicherweise Interaktion zwischen den Techniken), Warum genau diese Techniken, welche Zwecke erfüllen sie für die Anwender, Warum wurden andere nicht umgesetzt

## 4 Implementierung

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**

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**

### **5.2 Anwendung Visualisierung Zwei**

### **5.3 Anwendung Visualisierung Drei**

## **6 Verwandte Arbeiten**

Zwei Artikel mit ähnlichen Zielen diskutieren

## **7 Zusammenfassung und Ausblick**

Beiträge der Anwendung, welcher Mehrwert für Zielgruppe entsteht, mögliche Erweiterungen(Visualisierungen oder Daten)

## 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.

## Literatur

- [1] 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](https://www.tonight.de/games/fifa/fifa-ultimate-team-fut-wie-funktioniert-chemie-teamchemie-spielerchemie-links-loyalitaet_77330.html) (besucht am 07.09.2021).
- [2] *Min-Maxing*. de-DE. Feb. 2014. URL: <https://mein-mmo.de/lexikon/min-maxing/> (besucht am 04.09.2021).