

**Projektbericht zum Modul Information Retrieval und  
Visualisierung Sommersemester 2022**

## **Marktanalyse des Videospielsemarktes**

**Analyse und Visualisierung der Verkaufszahlen der Videospiele auf der Plattform  
XBoxOne**

Lena Arloth

4. Dezember 2022

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>3</b>
1.1	Anwendungshintergrund . . . . .	3
1.2	Zielgruppe . . . . .	4
1.3	Überblick und Beiträge . . . . .	5
<b>2</b>	<b>Daten</b>	<b>5</b>
2.1	Bereitstellung und Vorverarbeitung der Daten . . . . .	6
<b>3</b>	<b>Visualisierungen</b>	<b>7</b>
3.1	Analyse der Anwendungsaufgaben . . . . .	7
3.2	Anforderungen an die Visualisierungen . . . . .	10
3.3	Präsentation der Visualisierungen . . . . .	12
3.3.1	Visualisierung Eins . . . . .	12
3.3.2	Visualisierung Zwei . . . . .	13
3.3.3	Visualisierung Drei . . . . .	14
3.4	Interaktion . . . . .	15
<b>4</b>	<b>Implementierung</b>	<b>16</b>
4.1	Data.elm . . . . .	16
4.2	TreeHierarchy.elm . . . . .	17
4.3	ParallelPlot.elm . . . . .	18
4.4	Scatterplot.elm . . . . .	19
4.5	MainScatterParallel.elm . . . . .	20
<b>5</b>	<b>Anwendungsfälle</b>	<b>23</b>
5.1	Anwendung Visualisierung Eins . . . . .	23
5.2	Anwendung Visualisierung Zwei . . . . .	24
5.3	Anwendung Visualisierung Drei . . . . .	26
<b>6</b>	<b>Verwandte Arbeiten</b>	<b>27</b>
<b>7</b>	<b>Zusammenfassung und Ausblick</b>	<b>28</b>

# 1 Einleitung

Der weltweite Videospiegelmarkt über alle Plattformen hinweg ist mit einem Volumen von 156,17 Milliarden Euro im Jahre 2021 sehr groß und wird 2022 voraussichtlich auf 176,13 Milliarden Euro steigen.[4] Entsprechend viel Geld und Potenzial durch Wachstum verbirgt sich in ihm. Gleichzeitig benötigt es für die Entwicklung neuer Videospiele hohe Investitionen von Seiten der Entwicklerstudios und Verleger, folgend Publisher genannt, die erst nach einer gewissen Sicherheit durch Analyse des Marktes getroffen werden sollten. Um Prognosen und Tendenzen für die Zukunft finden zu können, braucht es die Erkenntnisse aus vergangenen Daten. Eine Möglichkeit hierzu ist die Analyse der totalen Verkaufszahlen der Videospieletitel aufgeschlüsselt nach Genre weltweit.

Als Problemstellung ergibt sich das fehlende Wissen der Entscheider innerhalb der Publisher, in welche Videospieletitel in welchem Genre und mit welcher Regionsausrichtung sie künftig investieren sollten, um das Potenzial des großen und wachsenden Videospiegelmarktes möglichst gut zu nutzen. Investition bezieht sich hierbei nicht zwingend auf eine konkrete Entscheidung für die Auftragsvergabe eines neuen Titels, sondern vor allem auf die Veranlassung detaillierter, umfassender Marktstudien. Erst dann kann eine fundierte endgültige Entscheidung getroffen werden. Es ergibt sich eine durchzuführende, grobe Marktanalyse der Verkaufszahlen der Videospiegelindustrie einer Plattform aus Sicht der Publisher. Sie fungiert als Grundlage und Unterstützung weiterer Entscheidungen und Investitionen.

Die Relevanz ergibt sich aus zuvor beschriebenem und der Notwendigkeit von Marktanalysen als Voraussetzung für unternehmerische Entscheidungen, Chancen und Risiken sowie Potenzialen.[2] Es ist für Unternehmen wie Publisher sehr wichtig zu erfahren, was und in welchem Genre künftig mit guten Verkaufschancen entwickelt werden soll. Weiterhin ist es relevant, mittels einer Marktanalyse zu erfahren, wie das Unternehmen selbst im Markt aufgestellt ist und was die Konkurrenz bietet. Um dies zu erreichen, ist es wichtig, Muster, Cluster und Korrelationen zwischen den Verkaufszahlen verschiedener Regionen in einem Genre zu erkennen. Für ein erleichtertes Verständnis sowie das Hervorheben von Auffälligkeiten sind Techniken aus dem Bereich der Visual Analytics von kritischer Relevanz.

Aus zuvor beschriebenem leitet sich folgende Fragestellung ab, die es mittels einer groben Marktanalyse aufbereitet, verdeutlicht und sichtbar gemacht durch Visualisierungstechniken zu beantworten gilt: In welches Genre und mit welchem Fokus auf die Regionen der Welt sollte ein Publisher künftig mit der Beauftragung detaillierter, teurerer Marktstudien zur Entscheidung über die Auftragsvergabe neuer Videospieletitel investieren?

## 1.1 Anwendungshintergrund

Marktanalysen sind für Unternehmen wie Publisher von zentralem Wert, da sie mit ihnen strategische Zukunftsfragen zur Ausrichtung beantworten können. Sie müssen unter anderem Fragen zu Produktverbesserungen, -erweiterungen oder -neuerungen, Investitionen in diese und Poten-

ziale und Risiken durch Korrelationen, Muster und Ausreißer beantworten können.[2] Dabei kann sie detaillierter oder oberflächlicher sein. In dieser Anwendung ist eine oberflächliche Analyse geplant, mittels derer Entscheidungen für detailliertere getroffen werden können. Der Vorteil in diesem Vorgehen liegt in der Ersparnis von Zeit und Geld.

Der Videospiegelmarkt ist in verschiedene Segmente geteilt, die wiederum unterteilt sind. Für diese Arbeit wird das Segment der Spieleplattform und Konsole *XBoxOne* von Microsoft gewählt. Dies liegt zum einen in der Verfügbarkeit der Daten begründet und zum anderen in der Beliebtheit der Konsole und der Konkurrenz zu anderen Plattformen wie *Playstation 4* und *Computer*. Gleichzeitig ergeben Analysen innerhalb einer Plattform für die Entscheidungsunterstützung künftiger Investitionen in Titel dieser Plattform mehr Sinn, da die Plattformen untereinander nicht kompatibel sind.

Die Konsole *XBoxOne* von Microsoft verkaufte sich besonders in den Jahren 2013 bis Anfang 2019 gut mit einer Spitze im Dezember 2015 mit rund 715.000 verkauften Einheiten im Monat. Mittlerweile ist die Beliebtheit der Konsole zwar stark zurückgegangen, doch lassen sich aus den Verkaufszahlen von Videospielen ihrer Plattform weiterhin Erkenntnisse für die Zukunft und potenzielle Investitionen der Publisher ableiten.[7] Hierbei geht es auch um Investitionen in neue Titel für die *XBoxOne*, aber vor allem um solche in die Nachfolgekonsolen *XBox Series X* und *XBox Series S*. Durch die Abwärtskompatibilität scheint es wahrscheinlich, dass Nutzer der *XBoxOne* auch die Nachfolgekonsolen nutzen.[5] Somit behalten die Verkaufszahlen der *XBoxOne* ihre, wenn auch leicht verminderte, Aussagekraft. Der Sinn einer überichtsartigen, grobgranularen Marktanalyse dieses Segments bleibt bestehen. Zur besseren Vergleichbarkeit und Aussagekraft der Verkaufszahlen und ihrer Auffälligkeiten der Videospiele ist eine Filterung nach Genre sinnvoll.

## 1.2 Zielgruppe

Die in dieser Arbeit betrachtete Zielgruppe setzt sich aus Entscheidern der Publisher von Videospielen zusammen. Diese Entscheider sind vorrangig Personen des oberen Management, aber auch des mittleren Managements, vorrangig aus dem Bereich Marketing und Forschung und Entwicklung. Das mittlere Management übernimmt dabei neben dem Mitentscheiden auch die Aufgabe des Präsentierens der Analysen mittels Visualisierungen für das obere Management. Weiterhin zählen in begrenztem Rahmen auch Stakeholder der Publisher zur Zielgruppe, da sie über Entscheidungen des Unternehmens zu künftigen Strategien informiert werden sollten.

In der Zielgruppe existiert kein detailliertes bzw. stark spezifiziertes Vorwissen zu speziellen Visualisierungstechniken. Jedoch ist anzunehmen, dass sich vor allem das mittlere und obere Management mit in der Betriebswirtschaftslehre häufiger vorkommenden Visualisierungen auskennt und diese ohne viele Erklärungen auswerten kann. Dazu zählen neben Balken- und Kreisdiagramme, Box Plots und Zeitreihendiagramme auch Scatterplots, explizite Baumdiagramme und Parallele Koordinaten Plots. Bei spezielleren Visualisierungstechniken wären mehr Erklärungen sowie mehr Zeit zur Analyse und Entscheidung nötig. Dies wäre für eine möglichst schnelle

und kostengünstige Entscheidung zu detaillierteren Marktstudien, wobei noch keine konkrete Entscheidung für einen Titel getroffen wird, nicht zielführend.

Durch die Visualisierungen werden mehrere Informationsbedürfnisse adressiert. Zum einen wird das Bedürfnis nach übersichtsartigen Informationen zur Videospielindustrie im Sinne der eigenen Position und der Konkurrenz für die *XBoxOne* bezüglich der Verbindungen von Publisher, Genre und Videospiel adressiert. Zum zweiten werden Informationsbedürfnisse nach Zusammenhängen und Mustern zwischen Regionen in den jeweiligen Genres befriedigt, um daraus strategische Entscheidungen ableiten zu können. Zum dritten wird das Bedürfnis nach konkreten Verkaufszahlen der einzelnen Spieletitel angesprochen.

### 1.3 Überblick und Beiträge

Die in dieser Arbeit verwendeten Daten stammen aus einem Datensatz für die *XBoxOne* mit den Attributen Tabellenposition, Publisher, Jahr der Veröffentlichung, Genre, Verkaufszahlen global sowie in den Regionen Nordamerika, Europa, Japan und Rest der Welt von der Plattform Kaggle.[6] Die Visualisierungen entwickeln sich vom Groben zum Detaillierten. In einer ersten Visualisierung wird zur Unterstützung einer Übersicht über die Publisher, Genre und Videospieltitel ein hierarchisches, explizites Baumdiagramm verwendet. Dadurch wird zudem ein Ansatzpunkt zur Auswahl der für die Publisher zu untersuchenden Genres gegeben. Zur Erkennung von Mustern und Ausreißern in den Verkaufszahlen eines Genres auch über alle Regionen hinweg bzw. in einer Ansicht aller Dimensionen gleichzeitig dient der folgende Parallele Koordinaten Plot. Zuletzt können in der dritten Visualisierung, dem Scatterplot, Korrelationen zwischen zwei Regionen sowie Ausreißer genauer betrachtet werden. In allen Visualisierungen sind Spieletitel mit Detailinformationen zu Publishern und Verkaufszahlen für einen konkreten Vergleich erkennbar. Mittels aller Visualisierungen kann die übergeordnete Fragestellung beantwortet werden, welches Genre durch positive Korrelationen, Tendenzen zu wenig Konkurrenz oder hohen Spielerzahlen durch viele Titel in dem Genre und der Analyse von Mustern und Ausreißern Potenzial bietet und damit detaillierterer Marktanalysen bedarf.

## 2 Daten

Der genutzte Datensatz "*Video Games Sales Dataset*" stammt von der Plattform *Kaggle* vom Nutzer *SID\_TWR*. [6] Die Daten entstanden laut *SID\_TWR* durch Erweiterung der Daten eines Web Scrapes von *VGChartz Video Games Sales* motiviert durch Gregpry Smith um weitere Attribute aus einem Web Scrape von *Metacritic*. Von den drei zur Verfügung gestellten Datensätzen, wird *XBoxOne\_GameSales* für dieses Projekt ausgewählt. Die zuvor erwähnten Erweiterungen um Attribute stammend von *Metacritic* sind hier nicht enthalten. Der Originaldatensatz liegt als CSV-Datei vor und beinhaltet zehn Spalten mit 613 einzelnen Positionen.

Zu jeder *Position* sind der Videospielname sowie das jeweilige Jahr der Veröffentlichung aufgelistet. *Genre* kategorisiert die Videospiele entsprechend und *Publisher* ordnet jedem Videospiel

seinen Verleger zu. Die Attribute *North America*, *Europe*, *Japan*, *Rest of World* und *Global* stellen die Verkäufe der Videospiele in *millions of units*, also Millionen Stück verkaufter Kopien, in diesen Regionen dar. Der für das Projekt gewählte Datensatz bildet Videospiele von 2013 bis circa 2020 ab. Aufgrund der fehlenden Aktualisierung der Verkaufszahlen auf *VGChartz Video Games Sales* ab dem Laufe des Jahres 2018, ist davon auszugehen, dass die für das Projekt verwendeten Daten kumulierte Verkaufszahlen ab 2013 bis circa 2018 abbilden.

Die vorhandenen Daten werden als gut geeignet für die Zielgruppe und das zuvor eingeleitete Zielproblem eingeschätzt. Sie ermöglichen eine grobe Übersicht über die verschiedenen Publisher, ihre Videospiele und bedienten Genres. Weiterhin schlüsseln sie detailliert auf, wie sich ein Videospiel kumuliert seit seiner Veröffentlichung global, aber auch in den einzelnen Weltregionen verkaufte.

## 2.1 Bereitstellung und Vorverarbeitung der Daten

Die Originaldaten werden als CSV-Datei heruntergeladen und zusätzlich zur Bearbeitung in *Open Office Calc* in das ODS-Format überführt. Alle Daten sowie das Projekt und dieser Bericht werden in einem öffentlichen GitHub Repository bereitgestellt. Die Datendateien sind im Ordner *Daten* in den Unterordnern *CSV*, *JSON* und *Tabelle* je nach Format zu finden.

Zur Erstellung des expliziten Baumdiagramms wird eine JSON-Datei benötigt, in der die Beziehungen der Publisher, Genre und Videospiele zueinander beschrieben sind. Videospiele seien Kinder der Genres, die wiederum Kinder der Publisher sind und unter einem Wurzelknoten zusammengefasst werden. Für die Realisierung des Scatterplots und des Parallelen Koordinaten Plots wird eine CSV-Datei benötigt. In dieser sind die nachfolgend detailliert beschriebenen Modifikationen enthalten.

Der Datensatz enthält fehlende Werte bei den Publishern und den Verkaufszahlen jeder Region sowie für die Anwendung nicht benötigte Informationen. Zur besseren Nachvollziehbarkeit der Datenvorverarbeitung wird im GitHub Repository im Ordner *Daten* eine zweite README.md-Datei erstellt, die den Fortlauf der Vorverarbeitung dokumentiert. Die benötigte Datengrundlage für vorliegende Arbeit wird in sieben Schritten erreicht, die folgend grob beschrieben sind und im Detail in angesprochener README nachvollzogen werden können.

Nach dem Herunterladen der Originaldateien und der Konvertierung dieser werden Testversionen der Originaldaten mit den ersten 20 Positionen im CSV- und ODS-Format erstellt. Sie werden um die Spalte des Erscheinungsjahres reduziert. Sie dienen der Übersichtlichkeit und Funktionalität in der Entwicklungsphase.

Es folgt die eigentliche Modifikation der Originaldaten für das Projekt. Dabei werden neue Dateien erstellt, um transparent die Originaldaten weiterhin zur Verfügung stellen zu können. Die Spalte des Erscheinungsjahres des jeweiligen Videospiele wird aus der hier durchgeführten Marktanalyse und ihrer Visualisierung ausgegliedert und gelöscht. Hieraus können weder Daten für ein Zeitreihendiagramm extrahiert werden, noch macht ein spezifischer Vergleich der Verkaufszahlen in Relation zum Erscheinungsjahr für die Problemstellung und deren Lösung

in einem ersten Überblick Sinn. Es werden zusätzlich jene Positionen, also Videospiele, eliminiert, die einen Wert von Null in den Verkaufszahlen aller Regionen bzw. automatisch global aufweisen. Sie sind ausgrunddessen fehlerhaft und haben in der Visualisierung keine Verwendung. Zur Ausschließung von Fehlern bei der manuellen Vorverarbeitung sowie universelleren Einsetzbarkeit des Quellcodes und Änderung der Rohdaten wird dieser Schritt zusätzlich in Elm programmiert. Aufgrund von Irrelevanz werden jene Positionen gelöscht, denen kein Publisher zugeordnet ist oder dessen Wert *Unknown*, sprich unbekannt ist. Zusätzlich wird der Datensatz um die Positionen minimiert, die von einem Publisher stammen, der nur ein Videospiel verlegt. Es ist davon auszugehen, dass sich jene Publisher sehr spezifisch ausgerichtet haben und keine hier visualisierte überichtsartige Marktanalyse benötigen. Weiterhin wird der Name des Publishers *Namco Bandai Games* in *Bandai Namco Games* vereinheitlicht, da sie durch Umbenennung 2014 denselben Verleger darstellen.

Mittels des Online-Tools *convertcsv.com* wird die Konvertierung der zusätzlich erstellten CSV-Datei in eine JSON-Datei vorgenommen. Diese wird auf Fehler in den Abhängigkeiten kontrolliert, der Wurzelknoten hinzugefügt sowie die entstandenen leeren Felder gelöscht. Zur Erzielung der gewünschten hierarchischen Struktur und Visualisierung wird pro Genre der jeweilige Eltern-Publisher wenn möglich sinnvoll abgekürzt in Klammern hinzugefügt. Auch hier wird eine Testdatei mit 20 Videospielen erstellt.

Eine Umrechnung der Einheiten der Verkäufe ist nicht ratsam, da sie schon übersichtlich vorliegt. Zur Erläuterung dieser dienen Informationstexte und die passenden Achsenbeschriftungen.

Die Herausfilterung der Positionen, die in nur einer Region Null-Werte in den Verkäufen aufzeigen wurde in Betracht gezogen, jedoch verworfen. Solche Videospiele würden nicht angezeigt, obwohl auch Informationen über keine erfolgten Verkäufe und die Region, in der nichts verkauft werden konnte, relevant sind. Zu beachten ist jedoch, dass als Null-Werte auch solche aufgeführt sind, deren Verkaufszahlen bei unter 0.01 Millionen Stück verkaufter Videospieldkopen liegen.

## 3 Visualisierungen

Im folgenden Kapitel wird eine Analyse der Anwendung zur Lösung des einleitend beschriebenen Zielproblems durchgeführt anhand derer Anforderungen an die Visualisierungen abgeleitet werden. Schließlich werden die Visualisierungen präsentiert.

### 3.1 Analyse der Anwendungsaufgaben

Das Zielproblem dieses Projekts ist eine überblicksartige, grobe Marktanalyse des Videospielemarktes mit Fokus auf den Verkaufszahlen der Plattform *XBoxOne* aus Sicht der Publisher. Ziel der Marktanalyse ist eine Übersicht über das Genre- und Videospielangebot des eigenen Verlagshauses sowie der Konkurrenz und Erkenntnisse über Zusammenhänge innerhalb von Genres in und zwischen verschiedenen Regionen der Welt durch Muster, Cluster und Korrelationen zu erhalten. Übergeordnet soll mittels dieser Informationen eine fundierte Entscheidung über

Investitionen in detailliertere, zielgerichtete Marktstudien in bestimmte Genres oder Regionen getroffen werden können, um den Publisher mit der Beauftragung neuer Videospiele strategisch für die Zukunft aufzustellen. Weiterhin sollen mit den Erkenntnissen erste Vorentscheide und Vorbereitungen für diese Spieleinvestitionen getroffen werden können.

Welche Publisher bieten welche und wieviele Videospiele in welchem Genre an? Gibt es Muster, Cluster, Ausreißer oder Korrelationen in den Verkaufszahlen zwischen den einzelnen Regionen der Welt oder gar über alle Regionen hinweg?

Durch eine schnell erstellte, intuitiv erklär-, präsentier- und verstehbare Visualisierung der grobgranularen Daten der Marktanalyse können im Publishingunternehmen erste vorab gültige Strategieentscheidungen bzgl. Investitionen in Titel eines Genres getroffen und Maßnahmen vorbereitet werden. Weiterhin kann dann eine detaillierte und deutlich aufwenigere und damit teurere Marktstudie in den zuvor identifizierten Genres und mitunter Regionen in Auftrag gegeben werden. Diese kann zu einer deutlich fundierteren Bestätigung oder Ablehnung der vorherigen Vorabentscheidung beitragen. Durch die Erstellung einer groben Marktanalyse und der Visualisierung derer Daten kann die Anzahl bzw. der Umfang detaillierterer, teurerer und zeitintensiverer Marktanalysen minimiert werden. Entsprechend sollte auch der Fokus einer Visualisierung auf Zeit- und Kosteneffizienz bei gutem Nutzen für zuvor beschriebenes liegen.

Wichtig dabei ist das Bedürfnis nach Intuitivität der Visualisierungen, Schnelligkeit im Verständnis und der Erstellung sowie des vergleichsweise geringen Preises für sie zu berücksichtigen. Die Visualisierungen müssen immer wieder schnell verständlich und erklärbar für andere sein, auch auf den ersten Blick.

Zunächst soll die Frage nach den angebotenen Videospiele der Publisher und ihrer Konkurrenz in einem ersten, qualitativen Überblick beantwortet werden. Der Fokus liegt aufgrund der gegebenen Daten und deren Übersichtsmöglichkeiten auf der Identifikation der von einem Publisher angebotenen Genres und der Videospiele, die jenen zugeordnet werden können. So lässt sich die vergangene Wahl des Schwerpunktes der Publishers bzgl. des Genres und der in diesem angebotenen Spiele erkennen. Besonders für größere Publisher mit vielen Videospiele oder vielfältigen Genres ist ein solcher Überblick sinnvoll. In Kombination mit einer Übersicht über die konkurrierenden Publisher in derselben Art, kann ermittelt werden, welcher Publisher eine starke Konkurrenz in welchem Genres darstellen kann, welche Genres stark oder weniger stark angeboten werden und wieviele Videospiele ihnen zugeordnet werden können. Somit können erste potenzielle Marktlücken, aber auch Sättigungen identifiziert werden, die für eine zukünftige Ausrichtung des Publishers relevant sein können. Es wird die Annahme getroffen, dass die Zielgruppe häufig und in unregelmäßigen Abständen einen Überblick über ihre Videospiele, deren Genres und ihre Konkurrenz auf den einen Blick benötigt. Dies kann sowohl im persönlichen Rahmen des Betrachtens durch Einzelpersonen des oberen und mittleren Managements der Fall sein als auch im Rahmen von Präsentationen der Analyse für das obere Management und wichtige Stakeholder. Durch die Übersicht soll bei der Zielgruppe ein mentales Modell der Hierarchie und Zugehörigkeit der Videospiele zu Genres und dieser Genres zu den Publishern entstehen.



Die zweite Unterfrage beschäftigt sich mit der Gewinnung von Erkenntnissen über Muster, Cluster, Ausreißer oder Korrelationen in den Verkaufszahlen zwischen den einzelnen Regionen der Welt und über alle Regionen hinweg. Dazu müssen die Daten so aufbereitet werden, dass sie in einer oder mehreren Darstellungen durch die Positionierung im Raum miteinander vergleichbar sind. Alle Erkenntnisse müssen schnell und einfach ersichtlich sein, sodass im Kopf der Personen der Zielgruppen leicht Bilder entstehen können. Datenpunkte in jedweder Form müssen sich zur Bildung von Mustern und Clustern zu Gruppen mittels Farbe, Position oder Form zusammenfinden können, die von anderen unterscheidbar sind. Ausreißer können ebenso erkannt werden. Korrelationen in den Verkaufszahlen zwischen den Regionen werden weiterführend mental meist am besten als Gruppe von Datenpunkten im Raum positioniert durch Absteigen, Aufsteigen oder Verstreuung erkannt.

Es ist schlussfolgernd sinnvoll, der Zielgruppe mittels einer Visualisierung erst eine Übersicht auf alle Regionen, sprich Attribute, auf einmal zu gewähren, in der Muster und Cluster sowie Ausreißer durch Markierung mit Farbe, Form oder Position über alle Regionen hinweg erkennbar sind. Mental soll ein Blick von oben auf die Welt entstehen. So kann sich ein zusammenhängendes Bild über die Verkaufszahlen in den Regionen ergeben ähnlich zu einem Motiv, dass sich durch das Verbinden von Zahlen in einem Zahlenbild für Kinder ergibt. Werden verschiedene solcher Bilder übereinandergelegt, hier entsprechend der konkurrierenden Videospiele, lassen sich Unterschiede erkennen. Sollte es zudem bspw. einen Ausreißer geben, der besonders gute Verkaufszahlen in dieser Region und global aufweist, dann deutet das auf eine Bedarfsdeckung dieser Region durch das entsprechende Videospiel hin. Eine weitere Investition in diese Region in dem Genre wäre entsprechend zu überdenken. Ist aber bswp. ein Videospiel des Publishers X in dieser Region der Ausreißer und besitzt auch global gute Verkaufszahlen, so ist eine Investition in weitere Videospiele dieser Reihe eine sinnvolle Option. In einem mentalen Modell können diese Analysen durch räumliche Abbildungen bzw. Abbildungen der Daten auf Positionen in einem Kontext dargestellt werden.

Mittels einer weiteren Visualisierung soll der Blick konkretisiert und auf zwei interessante Regionen fokussiert werden. Mental soll in das zuvor entstandene Bild hineingezoomt werden, um detaillierter auf zwei ausgesuchte Regionen des Genres blicken zu können. Sie können genauer auf Cluster und Ausreißer geprüft und die Art der möglichen Korrelation konkretisiert werden. Eine weitere Betrachtung der Gründe für letztere und der Ausnutzung dieser Korrelation in fortführenden Studien ist sinnvoll. Für die Bearbeitung dieser Unterfrage und vor allem der leichten Erkennung von Korrelationen und Clustern ist ein räumliches mentales Modell der Abbildung der Daten auf Positionen sehr wichtig und notwendig. Die Bestimmung der zu untersuchenden Regionen für weitere Investitionen in fortführenden Marktstudien kann so besser eingegrenzt werden.

Untergeordnet, jedoch sekundär auch für die Beantwortung der Hauptfrage wichtig, ist die Darstellung und Erkennbarkeit der Performanz der Videospiele des Genres in einer solchen Weise, dass einzelne Spiele korrekt erkannt und zugeordnet werden können. So können Ver-

kaufszahlen und damit Erfolge einzelner Spiele im Genre konkret verglichen werden. ufszahlen seiner Spiele feststellen. Bei der Bearbeitung dessen helfen dieselben mentale Modelle sowohl zur konkreten Ansicht als auch der gleichzeitigen Darstellung aller Attribute wie zur Beantwortung des zweiten Unterziels. Wichtig dabei ist die Möglichkeit der Identifizierung der Datenpunkte durch Beschriftung.

### **3.2 Anforderungen an die Visualisierungen**

Aus den zuvor analysierten Anwendungsaufgaben leiten sich Anforderungen an die Visualisierungen ab. Allgemein sollen alle Visualisierungen effizient, expressiv und angemessen sein. So ist bedingt durch die Zielgruppe und die Nutzung der Visualisierungen im Rahmen von Präsentationen vorrangig wichtig, dass die Inhalte intuitiv und ohne viele Erklärungen erkennbar werden. Weiterhin muss die Aussagekraft eindeutig sein, um unbeabsichtigte Suggestionen zu vermeiden. Zuletzt sollte besonders im Kontext einer betriebswirtschaftlich orientierten Analyse auch die Angemessenheit gegeben sein. Zwar ist eine hier anvisierte Marktanalyse essentiell für ein Verlagshaus und entsprechend auch bezüglich des Ressourcenverbrauchs wertvoll, jedoch werden besonders hier die Kosten in Bezug zur Effektivität gestellt. Es ist für eine grobe Marktanalyse wenig sinnvoll, sehr komplexe, tiefgründige Visualisierungen mit hohem Ressourcenverbrauch zu nutzen, wenn es besonders in Bezug auf die Zielgruppe sinnvollere, eventuell weniger komplexe, aber dafür leichter und schneller verständliche Visualisierungstechniken gibt, die entsprechend weniger Ressourcen in der Erstellung benötigen. Selbiges gilt für den Detaillierungsgrad der Daten, die für die Anwendungsaufgaben wie zuvor definiert grober sein dürfen und erst nach Entscheidung mithilfe der hier präsentierten Visualisierungen detaillierter und durch weitere Visualisierungstechniken komplexer dargestellt werden sollen.

Zur Beantwortung der einleitend ersten gestellten und zuvor analysierten Unterfrage ist eine überlicksartige, hierarchische Darstellung der Daten zu Publishern, den von ihnen bedienten Genres und in diesen angebotenen Videospielen notwendig. Die Darstellung sollte von den jeweiligen Publishern ausgehend spezifischer werden. Zur schnellen Erkennung der Videospieleangebote sollte nicht nur die Anzahl und Art der Genres je Publisher berücksichtigt werden. Aufgrund der mentalen Gewichtung und Bewertung der abgedeckten Genres muss auch die Darstellung der Videospieletitel je Genre des jeweiligen Publisher vorgenommen werden. Zur Ermittlung der Konkurrenz ist es nötig, alle Publisher in der Übersicht in derselben Art und Weise abzubilden. Neben diesen Anforderungen ist jedoch die schnelle Übersicht, vor allem über das eigene Verlagshaus die wichtigste. Es muss ohne viel Nachdenken und viele Erklärungen sowohl für die primäre Zielgruppe, als auch für die sekundäre, also Stakeholder des Verlages, ersichtlich werden, welche Genres der Publisher anbietet, welche und wieviele Spiele unter diesen Genres kategorisiert sind und wie die Konkurrenz aufgestellt ist. Die Hauptaufgabe der ersten Visualisierung liegt in der Präsentation der schon klar bestimmten Daten des Sachverhaltes der Publisher, der von ihnen abgedeckten Genres und den ihnen zugeordneten Videospielen in einer schnell verständlichen, hierarchischen Darstellung.

Aus der Analyse der zweiten Unterfrage ergibt sich die übergeordnete Anforderung der visuellen Analyse der Datenmenge für die zweite Visualisierung. Es muss in der Visualisierung in einem Vergleich auf einen Blick erkennbar werden, wie sich welche Spiele eines Genres in welchen Regionen der Welt verkauften. Somit ist es von Nöten, alle in den Daten gegebenen Regionen in einer Darstellung abzutragen, in der sie miteinander verglichen werden können. Die Visualisierung mehrdimensionaler Daten ist essentiell. Für eine bessere Vergleichbarkeit der Daten soll nach Genre kategorisiert werden und dieses je nach Betrachtungswunsch ausgewählt werden. Um die geforderten Muster und Cluster über alle Regionen hinweg zu visualisieren, ist eine visuelle Zuordnung zueinander mittels Farbe und Position wichtig. Muster sind bspw. gut durch Überlappungen der Datenpunkte erkennbar. Durch die exponierte Positionen können zudem Ausreißer erkannt werden. Es braucht eine deutliche Unterscheidung der Spiele untereinander, bspw. mittels Farbe, sowie eine Beschriftung mit den Titeln der Videospiele, um die untergeordnete genaue Erkennbarkeit aller Videospiele zu ermöglichen. Die Ausprägung der Verkaufszahlen muss zumindest grob ablesbar sein, wobei der visuelle Vergleich zur Unterstützung des mentalen Modells des übersichtsartigen Vergleichens vor expliziten, genauen Zahlen Vorrang hat.

Die übergeordnete Anforderung an die dritte Visualisierung ist wieder die visuelle Analyse der Datenmenge. Hierbei soll es der Zielgruppe möglich sein, im Detail Korrelationen zwischen den Regionen der Videospielverkäufe zu erkennen. So kann erkannt werden, wie die Zusammenhänge zwischen bspw. sehr guten Verkaufszahlen einer Region zu den Verkaufszahlen einer anderen gegeben sind. Weiterhin sollen noch einmal präziser die Verkaufszahlen der Videospiele der Publisher in einem Genre im Kontext dargestellt werden. Auch hier besteht die Anforderung, die Analyse und Visualisierung im Kontext des Genres vorzunehmen, damit die Zielgruppen die einzelnen Videospiele konkreter vergleichen können. In dieser Visualisierung muss die Relation von zwei Regionen zueinander dargestellt werden, sowie der Titel des entsprechenden Videospiel ersichtlich werden. Zum präzisen Vergleich von Verkaufszahlen in den Regionen müssen auch die exakten Zahlen zu finden sein und möglichst nicht nur rein visuell abgetragen werden. Für eine Visualisierung der möglichen Cluster und Zusammenhänge zwischen den Verkaufszahlen konkurrierender Spiele, müssen alle Titel aller Publisher des Genres vertreten sein. Durch die gewünschte detaillierte Analyse ist eine intuitiv verständliche Visualisierung sowie ein gut verständliches Design dieser wichtig.

Über alle Visualisierungen hinweg ist eine einheitliche Wahl der Farbgebung zur Verbesserung der Effektivität und Effizienz nötig. Um Probleme bei Sehschwächen der diversen Zielgruppe zu vermeiden, ist die Wahl nur einer Farbe zur Hervorhebung und Darstellung der Datenpunkte sinnvoll. Durch Abstufung mittels Deckkraft sind sie auch für Farbblinde erkennbar und verdeutlichen zudem bestimmte Erkenntnisse. Sogenannte *Just Noticable Differences*, also gerade so erkennbare Farb- und Kontrastunterschiede, in der Gestaltung der Website an sich helfen bei der Führung des Auges des Betrachters.

Allen spezifischen Anforderungen voran steht die Anforderung an die Angemessenheit. Die eingesetzten Visualisierungstechniken sollten günstig sein und deshalb einfacher gestaltet. Gleich-

zeitig sollten sie durch die Notwendigkeit von schnellen Entscheidungen zügig erstellbar und ebenso schnell und einfach verständlich sein, damit auch die in den Visual Analytics ungeübten Personen der Zielgruppe ein möglichst breites und präzises Verständnis für die Marktsituation, Verkaufszahlen und Korrelationen zwischen den Regionen erhalten können.

### 3.3 Präsentation der Visualisierungen

Zur Erfüllung der Anaylsen und Anforderungen werden die Visualisierungstechniken *explizites Baumdiagramm*, *parallele Koordinate* und *Scatterplot* gewählt. Nachfolgend werden die ausgetuchten Visualisierungen vorgestellt und diskutiert.

#### 3.3.1 Visualisierung Eins

Zur Erfüllung der Anforderungen wird für die erste Visualisierung ein explizites Baumdiagramm gewählt. Mit einem solchen können Hierarchien gut, übersichtlich und auf allgemein bekannte Weise dargestellt werden.

In der Visualisierung gibt es einen Wurzelknoten, der als Kindknoten die Publisher hat, die wiederum als Kindknoten die vom jeweiligen Publisher abgedeckten Genres besitzen. Zur besseren Übersicht und Vergleichbarkeit werden nur diejenigen Genres pro Publisher abgebildet, die dieser Publisher anbietet und in alphabetischer Reihenfolge aufgeführt. Wichtig dabei ist, dass keine sich überkreuzenden Pfade entstehen respektive es keinen Kindknoten gibt, der mehrere Elternknoten hat. Dies dient der Erfüllung der Anforderungen und der Übersichtlichkeit der Visualisierung. Die Genreknoten besitzen wiederum Kindknoten, die die Titel der Videospiele, die dem entsprechenden Genre zugeordnet sind, tragen. So entsteht die gewünschte hierarchische Übersicht über das Angebot jedes Publishers, welche durch die Beschriftung und Hervorhebung bei Hovern über jeden Knoten unterstützt wird. Die Anforderungen an eine schnelle, intuitive und für jeden verständliche Übersicht werden größtenteils erfüllt. Jedoch besteht der Nachteil der Größe des Baumes durch die vielen Publisher- und Genreknoten. Trotzdem kann der Baum durch seine Nutzung der Hierarchie, also Abbildung der nominalen Daten auf eine Position im Raum, sowie durch die Nutzung von Verbindungslinien zur Unterstützung der Positionierung und damit Erstellung des mentalen Modells der gewünschten Hierarchie die Anforderungen an Effektivität gut erfüllen. Entsprechende visuelle Variablen werden als sehr effektive Abbildungsmethodik für nominale Daten eingestuft. Expressiv und damit ohne viele Erklärungen und Missverständnisse ist diese Technik zudem. Jedoch leidet die Expressivität etwas unter der Menge an Knoten und damit der Größe des Baumes. Auch angemessen ist eine Visualisierung als explizites Baumdiagramm, da es wenig Ressourcen fordert.

Alternativen zu dieser Darstellung sind implizite Baumdiagramme, radiale Baumdarstellungen und hyperbolische Bäume. Implizite Baumdarstellungen können ebenso wie explizite die Hierarchien zwischen Eltern- und Kind-Knoten darstellen. Radiale und hyperbolische Bäume sind gute Alternativen für Darstellungen mit vielen Knoten wie es in dieser Visualisierung auch der Fall

ist. Vor allem letztere Alternativen bieten in Bezug auf Expressivität, Effektivität und Angemessenheit nahezu dieselben Vorteile wie die expliziten Bäume. Letztlich wird die Entscheidung für ein explizites Baumdiagramm trotz des Nachteils der vielen Knoten, die durch hyperbolische und radiale Bäume übersichtlicher gestaltet werden könnten, aufgrund der Übersichtlichkeit durch Intuitivität der expliziten Baumdarstellungen getroffen. Dies mag paradox wirken, jedoch sind explizite Bäume durch ihr häufigeres Vorkommen in allen Lebenssituationen noch ein wenig intuitiver und damit effektiver verständlich als die Alternativen. Da die Anforderung vor allem auf schnellem Erkennen und Wiedererkennen insbesondere des eigenen Verlagshauses liegt, wird diesem Aspekt eine hohe Gewichtung gegeben.

### 3.3.2 Visualisierung Zwei

Als zweite Visualisierungstechnik zur Teillösung der zweiten Untieranwendung werden die parallelen Koordinaten gewählt.

Für diese Visualisierung werden die Datenpunkte als Abfolge von Liniensegmenten dargestellt, wobei der Attributwert der jeweilige Endpunkt des Liniensegments ist. Visualisiert werden die Wertebereiche der Attribute *North America*, *Europe*, *Japan*, *Rest of World* und *Global*, welche die Verkaufszahlen in Millionen Einheiten der Videospiele in den verschiedenen Regionen darstellen. Durch die Verbindung der Liniensegmente und den jeweiligen Endpunkten auf den parallelen Liniensegmenten bzw. Achsen kann die Zielgruppe die Attributwerte dieser fünf Attribute ablesen. Durch die Darstellung aller Videospiele als je eine solche Sequenz von Liniensegmenten können Unterschiede erkannt werden.

Wie in den Anforderungen beschrieben, soll es zur besseren Vergleichbarkeit der Datenpunkte und damit Videospiele eine Filterung nach Genres geben, die mittels eines Drop-Down-Menüs umgesetzt wird. Da es bei parallelen Koordinaten zu Problemen mit dem Überzeichnen von Liniensegmenten kommen kann, wenn es zu viele Datenpunkte gibt, schafft die Filterung hierbei zudem Abhilfe.

Durch die Implementierung einer Art umgekehrten Röntgenstrahleneffekt durch die Nutzung einer Farbe mit wenig Deckkraft kann dieses Problem zusätzlich gelöst werden. Zudem können Überlappungen dadurch besser zur Visualisierung der geforderten Muster, aber vor allem Cluster genutzt werden. Je dunkler bestimmte Segmente, desto mehr Datenpunkte laufen hier überein. Auch Überkreuzungen von Datenpunkten sind einfach durch die dunklere Farbe am Kreuzungspunkt erkennbar.

Zur Erkennung des Videospieltitels und des Publishers, wird beim Hovern über ein Liniensegment beides eingeblendet. Die Verkaufszahlen sind anhand der Achsenbeschriftung ungefähr ablesbar, beim Hovern allerdings konkret. So überwiegt wie gewünscht das visuelle Erkennen und Vergleichen, aber auch die genaue Erkennbarkeit der Videospiele ist gegeben. Zudem wird der entsprechende Datenpunkt mittels Farbe hervorgehoben, um ihn eindeutig erkennbar zu machen. Dadurch können die konkreten Anforderungen sowie die allgemeinen Anforderungen an Expressivität und Effektivität an die zweite Visualisierung erfüllt werden.

Quantitative Daten werden weiterhin am besten mit den visuellen Variablen Position, aber auch Orientierung dargestellt. Beides wird in den parallelen Koordinaten umgesetzt, die Orientierung vor allem durch die Sequenz an verbundenen Liniensegmenten, die parallelen Achsen und den entstehenden Kontext der Daten.

Die Expressivität wird zusätzlich durch die eindeutige Beschriftung der Liniensegmente beim Hovern, die Einfärbung und die eindeutige Beschriftung der Achsen erreicht. Auch die weiteren Anforderungen vor allem an die Vergleichbarkeit und Erkennung von Mustern in einem Genre über alle Regionen hinweg in einer leicht verständlichen Art und Weise wird erfüllt.

Eine der Alternativen zu sind parallelen Koordinaten die Icon Techniken. Sie werden jedoch als weniger gut im Vergleich zu den parallelen Koordinaten eingeschätzt. Zuletzt bieten Sternkoordinaten und auch Polar Plots alternative Möglichkeiten der Darstellung mehrdimensionaler Daten. Allerdings

### 3.3.3 Visualisierung Drei

Zuletzt wird zur weiteren Teillösung der zweiten Unteranwendung die Visualisierungstechnik Scatterplot umgesetzt.

In diesem Scatterplot wird auf die X-Achse je nach Auswahlwunsch des Betrachters je der Wertebereich der Attribute *North America*, *Europe*, *Japan*, *Rest of World* und *Global* abgebildet. Die Abbildung der Wertebereiche auf die Y-Achse funktioniert auf dieselbe Art. So kann der Anwender frei zwischen zwei Attributen seiner Wahl wählen, die er untersuchen möchte. Es werden je die Datenpunkte visualisiert, die dem per Drop-Down-Menü ausgewählten Genre entsprechen.

Die Datenpunkte werden klassisch als ungefüllte Kreise dargestellt, wobei sich ihre Position aus den Werten der oben beschriebenen, momentan ausgewählten Attribute bestimmt. Beim Hovern über die Punkte werden diese eingefärbt und ein weiteres Attribut, der Titel des Videospiels angezeigt. Zur exakten Erkennung der Attributwerte der Verkaufszahlen werden diese hinter dem Titel in Klammern aufgelistet. Es wird zuerst der Wert für die X-Achse, dann der Wert für die Y-Achse dargestellt.

Durch diese Designentscheidungen werden die zuvor formulierten Anforderungen erfüllt. Die Kodierung der Daten auf die Punkte im Scatterplot, also Koordinatensystem ermöglichen ein Wahrnehmen und damit eine Orientierung im Raum, der mental entsteht und das Verständnis für die Relation der Datenpunkte zueinander verbessert. Weiterhin ist dadurch die visuelle Darstellung von möglichen Korrelationen zwischen zwei Attributen erkennbar.

Durch die zweidimensionale Darstellung der Attribute kann eine genauere Betrachtung dieser erfolgen. Entsprechend können Korrelationen besser analysiert werden. Auch die zuvor schon erkannten Muster und Ausreißer können überprüft werden. Durch die Einblendung der exakten Attributwerte ist auch die untergeordnete Anforderung der genauen Erkennbarkeit der Videospiele erfüllt.

Die eher klassische Darstellung als Punkte erleichtert das Erkennen von Punktwolken respek-

tive Cluster durch Intuition, insbesondere sollte es Korrelationen zwischen einer Region und den globalen Verkaufszahlen geben. Ähnlich wie in Visualisierung zwei werden auch in Scatterplots die für quantitative Daten gut geeigneten visuellen Variablen Position, Orientierung und Gebiet genutzt.

Weiterhin ist diese Darstellungsweise allgemein bekannt, sodass eine schnelle Analyse einzelner Wertpaare, aber auch Korrelationen und Ausreißer möglich ist. Dadurch und durch die oben genannten Aspekte ist die Effektivität gegeben. Auch expressiv ist diese Visualisierung, da die Darstellung der Datenpunkte als Kreise auch ohne Legende schnell erkannt wird. Durch die Einblendung der exakten Werte beim Hovern wird eine mögliche Verwirrung durch Überlagerung verschiedener Symbole, Farben oder Größen der Punkte verhindert. Die Wahl einer Farbe zur Darstellung der Datenpunkte, Füllung dieser beim Hovern und Wahl einer nicht einhundertprozentigen Deckkraft zur Ermöglichung der vereinfachten Erkennung der Punkte bei Überdeckung verbessert die Expressivität weiter. Durch die unkomplizierte und schnell erstellbare Darstellung ist auch die Angemessenheit gegeben.

Alternativ bestehen in diesem Kontext keine sinnvollen Visualisierungsmöglichkeiten. Q- und QQ-Plots stellen Attributwerte nur F-Werten gegenüber bzw. auch gegeneinander, dann jedoch um Verschiebungen zueinander zu erkennen. Korrelationstabellen lassen die hier vor allem gewünschten Korrelationen genau analysieren, stellen diese jedoch nicht visuell dar. So leiden Expressivität und Effektivität stark und die Anforderungen an Intuitivität, sowie Überprüfung von Mustern und Ausreißern sind nicht gegeben. und Korrelationstabellen.

### 3.4 Interaktion

In der ersten Visualisierung werden Knoten bei Interaktion durch Hovern über sie mittels Farbe hervorgehoben sowie beschriftet. So ist eine allgemeine Übersicht ohne zu viele Details auf einmal möglich. Die detaillierten Visualisierungen sind durch die Interaktion des Klicks auf die Weiterleitung erreichbar. So wird die Struktur der zuerst angebotenen Übersicht und der dann folgenden Detaillierung mittels Verkaufszahlen ermöglicht. In dem Parallelen Koordinaten Plot sind die Regionen pro Achse frei wählbar. Ein einfacher Achsentauch ist mit steigender Anzahl der Achsen unüberichtlich und beansprucht in der Auswahl zu viel Zeit. Die freie Wahl ermöglicht der Zielgruppe mehr Schnelligkeit und Einfachkeit in der Wahl der gewünschten Reihenfolge. Weiterhin werden die Datenpunkte bei Interaktion durch Hovern über sie mittels Farbe hervorgehoben und die Details außerhalb des Koordinatensystems angezeigt. Im Scatterplot wird ist die Belegung der X- und Y-Achse ebenso frei mit den vorhandenen Regionen belegbar. Die Zielgruppe kann somit individuell wählen, welche Regionen durch zuvor betrachtete Parallele Koordianten interessant sind und dargestellt werden. Auch hier werden bei Interaktion durch Hovern die Datenpunkte durch Farbe hervorgehoben sowie Details zum jeweiligen Videospiel angezeigt. Somit entsteht eine Einheitlichkeit in allen Visualisierungen, vor allem aber in der zweiten und dritten.

Weiterhin kann durch die Interaktion mit einem Dropdown-Menü zwischen allen drei Visual-

sierungen gewechselt werden. Änderungen der Achseneinstellungen im Parallelen Koordinaten Plot und im Scatterplot bleiben beim Wechsel zwischen diesen bestehen, sodass eine erneute Einstellung nicht nötig ist. So sind schnelle Vergleiche und Wechsel zwischen beiden detaillierten Visualisierung möglich. Das Baumdiagramm als Ausgangsübersicht ist zudem mittels Interaktion mit Link über jede der beiden anderen Visualisierungen möglich. In den Anforderungen für Visualisierung zwei (Parallele Koordianten) und drei (Scatterplot) ist eine Filterung nach Genre formuliert. Diese ist per Interaktion mit Dropdown-Menü und Auswahl eines Genres erreichbar. Wird dies in einer der beiden Visualisierungen ausgewählt, wird es in der anderen übernommen ohne bei Wechsel zu dieser Visualisierung nochmal ausgewählt werden zu müssen. Eine Änderung des Genres in einer der beiden detaillierten Visualisierungen führt zu einer Änderung in der anderen.

## 4 Implementierung

Im folgenden Kapitel wird die Implementierung der zuvor beschriebenen Visualisierungen und Interaktionen beschrieben. Der Code wird in fünf Elm-Module unterteilt. Es existiert je ein Modul für die drei einzelnen Visualisierungen (TreeHierarchy, ParallelPlot, Scatterplot), eines für Datenstrukturen (Data) und eines zur Zusammensetzung und Ausführung des Programms (MainScatterParallel). Innerhalb der Module, insbesondere der Visualisierungsmodule, wird versucht, den Code einheitlich zu strukturieren. Es werden zuerst kleinere Funktionen definiert, die im späteren Verlauf genutzt werden bis zuletzt allgemeine Einstellungen der Plots vorgenommen werden.

Über alle Module hinweg werden diverse Bibliotheken genutzt, die zu Beginn jedes Moduls importiert werden.

Insgesamt war die Implementierung der Anwendung sehr zeitaufwändig, da vor allem die Routine im Programmieren mit Elm trotz der umfangreichen Übungsaufgaben fehlte. Die Übungsaufgaben waren eine große Stütze. Jedoch musste auch hier noch einmal nachvollzogen werden, was dort gemacht wurde. Teils kamen so noch neue Erkenntnisse und neues Verständnis für den Code, sodass sich durch die Implementierung der Projektanwendung ein zusätzlicher großer Lerneffekt einstellte. Es gab trotz der guten Vorlagen der Abgaben der Autorin zu den Übungsaufgaben einige Schwierigkeiten, die in den folgenden Unterkapiteln Erwähnung finden.

### 4.1 Data.elm

In diesem Modul sind nahezu alle verwendeten Datentypendeklarationen enthalten. Einzig der *type Model* und *type Msg* aus Gründen der Übersichtlichkeit und Verständlichkeit des Codes in *MainScatterParallel* sind dort nicht vertreten.

In *Data* sind zunächst die Funktionen *decodeGameSales*, *csvString\_to\_data* und *gamesSalesList* zur Decodierung der CSV-Dateien implementiert. Die einzelnen Felder der CSV-Datei als String decodiert und dann in eine Liste vom Typ *GameSales* geschrieben. Darauf folgt die Deco-



dierung der JSON-Datei mittels *treeDecoder*. Beide Decoder basieren auf jenen für die Übungen sieben und neun verwendeten und wurden an die in diesem Projekt verwendeten Daten angepasst.

Es folgt die Definition der Datentypendeklarationen, die für alle Visualisierungen außer die *TreeHierarchy* wichtig sind. In *type alias GameSales* sind die Daten aus der CSV-Datei definiert, welche dann im Model in *MainScatterParallel* weiterverwendet werden. Die Deklarationen *type RegionType* und *type PlotType* als CustomType werden für die Auswahl der Achsen in den Visualisierung zwei und drei sowie die Auswahl der anzuzeigenden Visualisierung benötigt. Sie werden im *type Model* und im *type Msg* in *MainScatterParallel*, im *regionFilter* in *Scatterplot* und im *multiDimenData* in *ParallelPlot* verwendet. Als nächstes werden die spezifischen Datentypen für die Visualisierungen definiert.

Zuletzt werden die für die Buttons wichtigen Funktionen zur Datentypkonversion von einem String zu PlotType, String zu RegionType, RegionType zu String und String zu ein einem Tupel aus RegionType und String beschrieben. Dies ist nicht nur zur Darstellung von Buttons, die nur Strings erlauben, also der Implementierung von Interaktionen nötig, sondern auch zur bspw. korrekten Anzeige von aktuell ausgewählten Regionen. Weiterhin ist eine Konversion von einem RegionType zu einem Tupel aus GameSales und Float für die Funktion *multiDimenData* in *ParallelPlot* notwendig.

## 4.2 TreeHierarchy.elm

Nach dem Import der Bibliotheken beginnt dieses Modul mit den grundlegenden Funktionen *main*, *init* und *update*. Die Anwendung wird initialisiert, die Daten unter Anwendung des Decoders für JSON-Dateien geladen und die jeweils auszuführenden Aktionen bei korrektem oder inkorrektem Laden der Daten bestimmt. Es folgt eine Helferfunktionen *convert* zur Konvertierung des Baumes basierend auf den Vorgaben zu Übung neun.

Die Funktionen *line* und *point* werden zum Zeichnen der Verbindungen von Eltern zu Kindknoten bzw. zur Darstellung der Knoten an sich benötigt und in *treePlot* und *treePlot2* angewandt. Sie definieren die Linie bzw. den Knoten als Kreis mit Beschriftung mit einer Rotation von 75° und einer fett gedruckten Serifenschriftart.

Als nächstes sind zwei Funktionen zum Zeichnen des Baumes beschrieben. Zur Ansicht des Baumes ohne Scrollen als Übersicht wird *treePlot2* verwendet, zur näheren Ansicht des Baumes *treePlot*. Die Plots unterscheidet einzig die prozentuale Vergrößerung der global definierten Höhen und Breiten um 200 bzw. 150 Prozent im SVG des *treePlot*. Die Plots an sich basierend stark auf dem Code der Autorin zu Übung neun. Mittels lokaler Funktionen wird zunächst das Layout des Baumes bestimmt, bevor lokal die Abhängigkeiten des Baumes berechnet werden, um die Pfade zwischen Eltern- und Kindknoten zu zeichnen. Dazu werden die X- und Y-Werte der Eltern- und Kindknoten mittels *Dict.get* herausgezogen. Weiterhin wird eine Verbindung des Wurzelknotens mit einem nicht existierenden Elternknoten durch die Funktionen *checkRootNegative* und *nodeValuePath* verhindert. Zum Zeichnen der Verbindungen wird die zuvor

beschriebene Funktion *line* genutzt, auf die die lokale Funktion *nodeValuePath* angewandt wird. Für die Punkte wird *point* verwendet, auf das die lokale Funktion *nodeValue* angewandt wird. Die Definition von Kreis-, Linien- und Textfarben, Umrandungen derselben und entsprechenden Änderungen beim Hovern mit der Maus wird im global definierten CSS *cssTree* vorgenommen.

In der *view*-Funktion wird lokal der konvertierte Baum sowie das Layout berechnet. In einem Div-Container werden geschachtelt durch weitere Container Überschriften, Erklärungen, Links und Buttons zur Ausgabe implementiert. Die Anordnung sowie der Designstil ist an *MainScatterParallel* angepasst.

Zuletzt werden global allgemeine konstante Einstellungen zum Plot definiert. Hierzu zählen Breite, Höhe, Innenabstand, Radius, Skalierungen, Default-Einstellungen sowie Einstellungen zum Bereich der Datenwerte. Diese Einstellungen basieren auf den Vorgaben in der Vorlage zum Arbeiten mit Bäumen aus Übung neun und sind auch in den Abgaben zu Übung neun der Autorin zu finden.

Die Implementierung der expliziten Baumhierarchie bereitete wenig Schwierigkeit, da sehr viel auf den abgegebenen Codes der Autorin zu Übung neun basiert. Lediglich der Decoder in Data musste an die Datenfelder angepasst, der Link zum Laden der Daten geändert sowie die besten Design- und Darstellungsweisen, inklusive der zweifachen Darstellung des Baumes, gefunden werden. Weiterhin wurde das CSS global definiert, das allgemeine Design der darzustellenden Seite im *view* an jenes aus *MainScatterParallel* angepasst und Erklärungstexte für die Anwender hinzugefügt.

### 4.3 ParallelPlot.elm

Dieses Modul beginnt nach dem Import mit drei Funktionen zum Mapping und der Reduzierung der Datensätze um solche, die Null-Values enthalten. Dies ist nach der manuellen Vorverarbeitung der Daten zwar nicht zwingend notwendig, jedoch werden somit mögliche Fehler in dieser berücksichtigt und entfernt. Weiterhin ist der Code damit universeller einsetzbar, sollten andere Datengrundlagen genutzt werden wollen.

Die erste Funktion ist eine Hilfsfunktion, die mittels *Maybe.map2* beim Pipen hilft. Die zweite Hilfsfunktion basiert auf einer normalen *Maybe.map5* Funktion, die um einen Parameter erweitert wird. Beide basieren auf den Dokumentationen der Elm-Core-Bibliothek zu Maybe. Da es kein *Maybe.map6* gibt, wird hier die vorherige Hilfsfunktion angewandt, um dieses Problem zu umgehen. *AssignmentAndReduce* wiederum wandelt schließlich lokal mithilfe der *helpMapBig* die Daten vom Typ GameSales zu Maybe GameSales um und reduziert diese um die Datensätze mit fehlenden Werten mittels *List.filterMap*. Diese Funktion basiert auf den Abgaben der Autorin zu Übung sechs.

Die Funktion *multiDimenData* dient zusammen mit der lokalen Funktion *multiDimFunction* aus *MainScatterParallel* der freien Auswahl und Belegung der Achsen sowie der Anwendung der Filterung nach Genre. Das Grundgerüst beider Funktionen basiert auf den Abgaben der Autorin zu Übung sechs. Der dort implementierte Achsentauch wurde aufgrund der Anforderungen in

diesem Projekt geändert. *MultiDimenData* benötigt als Eingabe eine Liste von GameSales sowie fünf RegionTypes als die auszuwählenden Regionen für die Achsen. Weiterhin braucht es den Namen des Spieles und den Publisher des Spieles zur Anzeige dieser im späteren Plot. Der Typ begründet sich aus dem *type alias GameSales*. Zusätzlich sind fünf Namen, also Beschriftungen als String nötig. Begründet ist dies in der Definition der Recordfelder in *type Model* sowie dem für das *update* benötigten Datentypen (*RegionType, String*). Die Funktion muss den Typ *MultiDimData* zur Übergabe an *scatterplotParallel* und Zeichnung des Plots ausgeben. Eine besondere Schwierigkeit bei der Konstruktion der Funktion lag im Umfang dieser und im Verständnis des benötigten Datentyps. So muss der RegionType innerhalb der Funktion mittels *regionTypeToAxisAnnotation* nochmals umgeschrieben werden. Die Funktion wird in *MainScatterParallel* in der lokalen Funktion *multiDimFunction* angewandt und bekommt dort die von ihr benötigten Daten übergeben.

Es folgt der *scatterplotParallel*, der nahezu unverändert aus den Abgaben der Autorin zu Übung sechs übernommen werden kann. Hier werden unter anderem mittels lokaler Funktionen die Achsen in X-Richtung positioniert sowie die Achsen und ihre Beschriftungen in einem umgebenden Rechteck gezeichnet. Die Zeichnung der Datenpunkte wird mittels *Shape.line* und *Shape.lineaerCurve* realisiert. Zudem wird der beim Hovern über einen Datenpunkt anzuzeigende Text lokal beschrieben bevor diese Funktion dann zuletzt mit den entsprechenden Daten angewandt wird.

Auch hier wird das CSS global in *cssParallel* definiert. Ein besonderes Augenmerk liegt dabei auf der Definition der Opacity mit 0.5. So entsteht trotz weiß gefärbtem Umgebungsrechteck der gewünschte Röntgeneffekt. Mögliche Cluster und Auffälligkeiten in den Datenpunkten sind durch Überlagerungen so besser erkennbar. Durch das Einfärben in einem stärkeren Grünton beim Hovern und dem Einblenden der Details zum Datenpunkt lassen sich die einzelnen Videospiele besser nachvollziehen.

Auch hier schließt die Implementierung mit den generellen Einstellungen für den Plot. Diese basieren auf den Abgaben der Autorin zu Übung sechs.

Die Erstellung des ParallelPlot basierte in weiten Bereichen auf den Abgaben zu Übung sechs. So bereitete das Zeichnen des Plots wenig Schwierigkeiten. Etwas komplizierter war zunächst die Implementierung des Mappings und der Filterung nach Null-Werten durch das fehlende *List.map6*. Schwieriger gestaltete sich jedoch die Umsetzung der freien Auswahl der Achsen wie oben schon beschrieben. Dies war mit viel Ausprobieren und Verständnis, welcher Datentyp für welche weitere Funktion benötigt wird und welche durch die vorherige Definition der Datentypen vorhanden waren, verbunden.

#### 4.4 Scatterplot.elm

Nach dem Import der Bibliotheken folgen in diesem Modul ähnliche drei Funktionen mit demselben Zweck wie im *ParallelPlot*. Die ersten beiden Funktionen sind dieselben. *FilterAndReduceGames* ähnelt *AssignmentAndReduce* stark, basiert aber auf den Abgaben der Autorin zu

Übung eins. Hier werden die Daten zunächst mithilfe von *helpMapBig* vom Typ *GameSales* dem Typ *Maybe Point* zugeordnet, welcher für die Funktion *point* und diese für den Plot an sich benötigt wird. Die zugeordneten und mit *List.filterMap* gefilterten Daten in *filter* werden dann unter anderem in *XyData* geschrieben, dass als Output der Funktion *FilterAndReduceGames* definiert ist. *XyData* werden benötigt für das Zeichnen des Scatterplots, da sie unter anderem eine Liste Daten des Typs *Point* enthalten.

Die Funktion *regionFilter* filtert mit Eingabe einer Liste *GameSales* und des *RegionTypes* die Datenpunkte nach den je gewünschten Regionsauswahlen für die Achsen. Diese Funktion ist nötig zur Interaktion innerhalb des Plots, um die Achsen frei wählen und die richtigen Datenpunkte anzeigen zu können. In *MainScatterParallel* wird *regionFilter* lokal für *valuesX* und *valuesY* aufgerufen und auf die nach Genre gefilterten Daten sowie die X- bzw. Y-Achse angewandt. Die Funktion basiert auf den Abgaben der Autorin zu Übung vier.

In der Funktion *point* wird die Positionierung der Punkte sowie deren Beschriftung für die Weiterverwendung in *Scatterplot* festgelegt. Der *scatterplot* ist für das Zeichnen des Plots mit den Datenpunkten zuständig. Mit der lokalen Funktion *pointsXY* wird die Variabilität der Achsen und damit X- und Y-Werte der Punkte ermöglicht, die für die Interaktivität der Achsenauswahl nötig ist. Dies ist in vielen Übungsabgaben der Autorin nicht umgesetzt, jedoch in Übung vier, sodass diese als Grundlage für die lokale Funktion dient. Im SVG des *scatterplot* wird die Positionierung der X- und Y-Achse, ihrer Beschriftungen und der Punkte festgelegt. Letzteres wird durch die lokalen Funktionen *xScaleLocal*, *yScaleLocal* und *pointsXY* sowie der zuvor beschriebenen globalen Funktion *point* erreicht. Der *scatterplot* basiert auf den Abgaben der Autorin zu Übung 1.4 sowie Übung vier.

Zuletzt werden allgemeine, konstant für den Plot geltende Einstellungen global definiert. Diese basieren auf den Abgaben der Autorin zu Übung eins bis vier.

Die Implementierung des Scatterplots verlief in weiten Teilen problemlos, wenn auch nicht schnell durch ständiges Überprüfen der richtigen Datentypen. Schwieriger war wie beim *ParallelPlot* auch die Umsetzung des Mappings und Filterns nach Null-Werten durch das fehlende *List.map6*. Weiterhin musste beachtet werden, dass die Achsen des Plots frei wählbar, also variabel definiert sind und nicht strikt festgelegt wie in den meisten Übungsserien. Mit entsprechender Zeit- und Denkinvestition war auch dies überwindbar.

## 4.5 MainScatterParallel.elm

Auch das Modul *MainScatterParallel* definiert zunächst die Importe. Aufgrund der besseren Erkennbarkeit aus welchen Modulen welche Funktionen stammen, werden keine Funktionen dezidiert importiert, sondern nur das allgemeine Modul.

Für dieses Modul wird die gesamte Elm-Architektur verwendet, da es den *Scatterplot* und die *ParallelPlot* ausführen und alle drei Visualisierungen verbinden soll. Hierzu wird das Programm zunächst wieder in der *main*-Funktion definiert, sowie die *subscriptions*- und die *init*-Funktion implementiert.

Danach folgt die Definition der Datentypendeklarationen *type Model* und *type Msg*. In *type Model* werden die drei Varianten Error, Loading und Success beschrieben, die in den entsprechenden Fällen angezeigt werden. Die Variante Success ist gleichzeitig ein Record aus den dort beschriebenen Feldern. Wichtig zu beachten ist, dass die initialen Werte erst in der *update*-Funktion hier hinein geschrieben werden. Im Feld *data* können die geladenen und decodierten Daten als Liste vom Typ *GameSales* gespeichert werden. Zur Ermöglichung der Interaktionen gibt es weiterhin Felder für das Genre, die Achsen eins bis fünf, die Namen der Achsen eins bis fünf, die X- und Y-Achse und den gewünschten Plot. Für die Achsen des Scatterplots und Parallelen Koordinaten Plots sowie den Plot an sich sind die entsprechenden CustomTypes definiert. Der *type Msg* definiert die Varianten für die Nutzung in *update* zur Überschreibung des Models. Es gibt eine Variante zum erfolgreichen Laden der Daten und damit der Initialisierung des Plots sowie für jeden möglichen Wechsel von Achsen in den Plots sowie den Wechsel des Plots an sich.

In der *update*-Funktion werden zunächst die initial anzuzeigenden Attributwerte definiert und in Success gespeichert. Hierbei muss auf die entsprechenden Datentypen geachtet werden, die in *type Model* unter Success gespeichert werden. Außerdem wird hier die Decodierung der geladenen Daten vorgenommen. Weiterhin werden für alle möglichen Interaktionen mit dem Model die Varianten von *Msg* eingefügt und beschrieben, wie bei einer Änderung dieser Variante durch Klick auf deinen Button das Model überschrieben wird.

Als nächstes werden alle benötigten Buttons als Funktionen implementiert, sodass das Model bei Klick auf eine Variante überschrieben werden kann. Da die Buttons nur Strings akzeptieren, die Varianten von *Msg* jedoch die CustomTypes bzw. die Tupel aus CustomType und String, muss hier bei allen Buttons außer dem *buttonGenreType* die entsprechende in *Data* definierte Umwandlungsfunktion genutzt werden. Die Änderungen werden dann in der Variante des *Msg*-Typen gespeichert und durch die *update*-Funktion das Model überschrieben. Die Buttons basieren auf den Abgaben der Autorin aus Übung vier.

Die Funktion *filterGenre* beruht auf den Abgaben der Autorin zu Übung vier und wird folgend im *view* angewandt. Mit dieser Funktion kann der Anwender nach dem Attribut Genre filtern und ein für ihn sehenswertes Genre auswählen.

In der *view*-Funktion werden die verschiedenen Fälle des Models behandelt. Für diese Erklärung wird nur auf den Fall des Erfolges beim Laden eingegangen. Zunächst wird eine *let-in*-Konstruktion auf erster Ebene erstellt. In lokalen Funktionen wird *gameSalesData* sowie die Anzahl der Spiele insgesamt beschrieben. Zusätzlich wird auf dieser oberen Ebene der globale *filterGenre* zur Filterung der Daten nach gewünschtem Genre angewandt. Weiterhin wird die Länge dieser neuen Liste, also die Anzahl der Spiele im ausgewählten Genre berechnet. Damit nun zwischen den Plots gewechselt werden kann, werden die verschiedenen Varianten bzw. Fälle des Datentyps *PlotType* im *in*-Teil angelegt. Deshalb befindet sich auch das Feld *plot* im Record der Variante *Success* des Models. Je nachdem, welcher Plot mittels Button ausgewählt wird und wie das Model durch das *update* überschrieben wurde, wird einer der drei Fälle angezeigt. Der Zustand des Models mit dem gefilterten Genre soll bestehen bleiben, sodass eine Interaktion

mit dem Genrebutton in einer Visualisierung auch in der anderen angewandt und übernommen wird. Dazu dient zum einen die zuvor beschriebene Variantendefinition im *in*-Teil der äußeren *let-in*-Konstruktion sowie jeweils eine weitere *let-in*-Konstruktion in den jeweiligen Falldefinitionen. Diese befinden sich auf zweiter bzw. innerer Ebene, sodass die in der äußeren Konstruktion berechneten Zustände, also Genreauswahlen, auch hier gelten und bei einem Wechsel des Plots übernommen werden. Um dieses Ziel zu erreichen müssen in den lokalen Funktionen der inneren Konstruktion der Plots die lokalen Funktionen der äußeren Ebene angewandt werden. Würde in jeder der inneren Ebenen für jeden Plot die Filter nach Genre definiert werden und lokal angewandt statt in der übergeordneten Ebene, würden die Einstellungen respektive der Zustand des Modells nicht übernommen werden.

Im ersten Case wird der ParallelPlot behandelt. Lokal wird nun die *assignmentAndReduce*-Funktion auf die gesamten Daten sowie zu in der äußeren Ebene nach Genre gefilterten Daten angewandt. In *multiDimFunction* wird *multiDimenData* letztlich genutzt und die entsprechenden Daten dafür übergeben. Als erste Liste vom Type *GameSales* wird die von Null-Werten bereinigte und nach Genre gefilterte *clearedGameSalesData* genutzt. Im *in*-Teil der inneren Ebene wird das Aussehen der Seite festgelegt. Hierbei wird darauf geachtet, dass zur angenehmeren und weniger ablenkenden Ansicht die Farben den Grüntönen in den Plots angepasst werden. Zur Direktion des Auges über die Seite und verbesserten Erkennung von wichtigen Abschnitten wie der Auswahl der Plots, des Genres und der Achsen werden Ränder und verschiedene Grünabstufungen genutzt. Neben Informationstexten und Überschriften werden hier auch die Buttons als Dropdown-Menü zum Wechsel des Plots, zur Änderung des Genrefilters sowie der Anpassung der Achsen eingefügt. Zudem wird transparent angezeigt, nach welchem Schritt wieviele Videospiele in der Auswahl verbleiben und welche Einstellungen gerade aktuell getroffen sind. Nach Anzeige der ausgewählten Einstellungen wird der Parallele Koordinaten Plot implementiert. Hierzu wird dem *scatterplotParallel* das *cssParallel* übergeben sowie zwei Konstanten für die Höhe und die Aspect Ratio. Schließlich wird die zuvor lokal berechnete *multiDimFunction* übergeben. Zuletzt wird noch ein Text mit Link direkt zurück zur TreeHierarchy angeboten.

Im zweiten Case wird der Scatterplot behandelt. Das Vorgehen mit den lokalen Funktionen zur Anwendung der Filter und des Mappings ist ähnlich wie zuvor. Durch die lokalen Funktionen *valuesX* und *valuesY* werden dem *regionFilter* die nach Genre gefilterten Daten als Liste von *GameSales* sowie die RegionTypes der X und Y-Achse übergeben. Somit wird der Filter nach Regionen angewandt und die Interaktion mit freier Auswahl der Achsen kann funktionieren. Der *in*-Teil der Plotvariante Scatterplot ähnelt dem des ParallelPlots stark und unterscheidet sich nur in einigen Informationstexten sowie der Auswahl der Achsen. Weiterhin wird hier dem *scatterplot* der *cssPoint*, die zuvor berechneten *gameSalesDataCleared* sowie die *valuesX* und *valuesY* übergeben. Um die Beschriftung der Achsen zu ermöglichen, werden diese mittels der Umrechnung von RegionType zu String als String hinzugefügt.

Da die TreeHierarchy in *TreeHierarchy.elm* implementiert ist, muss für diese Plotvariante nur das Html inklusive Link zum eigentlichen Plot des Baumdiagramms im selben Design wie zuvor

definiert werden.

Durch die zuvor in getrennten Anwendungen implementierten Visualisierungen war ein grundlegendes Zusammenfügen dieser durch stückweises Kopieren eher einfach umsetzbar. Besonders die Erstellung einer einheitlichen *update*-Funktion war zeitlich jedoch durch die Menge an Interaktionsmöglichkeiten aufwändig. Weiterhin schwieriger war die Implementierung der Buttons für den ParallelPlot, die zunächst je einzeln und nicht als Dropdown-Menü wie gewünscht vorlagen. Durch einige Denkarbeit, Ausprobieren vor allem mit der Funktion *multiDimenData* und verschiedenen Konvertierungsfunktionen konnten sie doch wie gewünscht realisiert werden. Auch hier musste zudem zur Überprüfung immer wieder zeitaufwendig das *update* angepasst werden. Zunächst wurden beide Plots so implementiert, dass sie untereinander gezeichnet wurden und es noch keinen Selektor für den Plot gab. Entsprechend war nur eine *let-in*-Konstruktion nötig und die Filterung wurde für beide Plots übernommen, was jedoch nicht dem Gewünschten entsprach. Nach schrittweisem Hinzufügen der Auswahl des Plots und allen zugehörigen Code-teilen, wurde durch Ausprobieren entdeckt, dass eine Übernahme des Modelzustands durch die Schachtelung der *let-in*-Konstruktionen im *view* und der Anwendung des Variantenwechsels für den Plot mittels *PlotType* möglich war. Hier war es im Verlauf kompliziert, den Überblick über die Schachtelungen sowie außen und innen definierten lokalen Funktionen zu behalten.

## 5 Anwendungsfälle

Im folgenden Kapitel wird ein möglicher Anwendungsfall der implementierten Visualisierungen dargestellt. Dabei werden diese aus der Sicht eines mittleren Managers des Publishers *505 Games* betrachtet, der sie und die Erkenntnisse aus ihnen dem oberen Management des Publishers sowie Stakeholdern des Unternehmens präsentiert. Daraus sollen Indikatoren für potentielle Investitionen in neue Videospiele bzw. Fortsetzungen vorhandener Titel eines Genres gefunden werden. Kann ein solcher Indikator für Potenzial gefunden werden, so kann entschieden werden, ob weitere detailliertere Marktstudien in Auftrag gegeben werden sollen. Zudem können Vorentscheidungen zur Aufstellung eines neuen Videospieles im Markt getroffen werden, wobei zwischen einer breiten, globalen Aufstellung oder der Konzentration auf bestimmte Märkte resp. Regionen entschieden werden kann.

### 5.1 Anwendung Visualisierung Eins

Zuerst lädt das explizite Baumdiagramm. Mittels farblicher Unterschiede der einzelnen Websitebereiche sowie der Abgrenzung der Container durch Umrandungen wird das Auge der Betrachter schnell und unbemerkt auf die wichtigen Steuerungselemente sowie die eigentliche Visualisierung gelenkt. Zu diesem Zweck wird mit Ausnahme des Navigationscontainers mit *Just Noticable Differences* gearbeitet. Sollte eine Person mit Farbsehschwäche unter den Präsentationsteilnehmern sitzen, so ist dies durch Abstufungen der Grüntöne auch im Kontrast auf der Website und in den Visualisierungen kein Problem. Falsche Farberkennungen spielen für das Verständnis keine

Rolle.

Im ersten Ausschnitt des Baumdiagramms kann der Präsentator den Teilnehmern einen Überblick über die Publisher gewähren sowie die eigene Position im Baum aufzeigen. Im zweiten, detaillierteren Ausschnitt wird den Managern und Stakeholdern dargestellt, welche Genres sie mit *505 Games* bedienen sowie die darunter kategorisierten Spiele. Insgesamt bieten sie fünfzehn Videospiele in acht Genres an, wobei *Adventure*, *Racing* und *Shooter* die meisten Spiele haben. Dank der Einblendungen der Titel und Genres beim Hovern über die sich grün färbenden Knoten in beiden Diagrammen sowie das Scrollen im zweiten Diagramm sind konkrete Ein- und Überblicke effektiv möglich. Es wird entschieden, die Genres *Action*, *Platform* und *Sports* aufgrund der bisher eher geringen Ausprägung im eigenen Unternehmen näher zu betrachten.

Aus dem Baumdiagramm ergibt sich durch die hohe Anzahl an anbietenden Publishern vor allem eine Konkurrenz im Genre *Action*. Durch die Struktur und Positionierung der Knoten ist schnell erkennbar, dass hier vor allem *Activision*, *Capcom*, *Ubisoft* und *Warner Bros.* durch ein hohes Spieleangebot die stärkste Konkurrenz zu sein scheinen. In diesem Genre kann Potenzial durch viele Käufer liegen, da viele Publisher hier Spiele anbieten. Gleichzeitig gibt es viel Konkurrenz, wodurch eine Sicht auf mögliche Korrelationen in den Regionen der Welt und die entsprechend beste Positionierung im Markt wichtig ist. *Sports* und vor allem *Platform* haben weniger Konkurrenz, wobei gerade ersteres durch Publisher wie *EA Sports* und *2K Sports* dominiert wird.

Aufgrund dieser ersten Übersicht entscheidet der Präsentator, dass nachfolgend der Fokus auf die Genres *Action* und *Sports* gelegt wird. Mit Klick auf den Link in der Navigation gelangt der Betrachter zu den detaillierteren Visualisierungen.

Die Informationen hätten auch durch ein implizites Baumdiagramm ermittelt werden können. Aufwand und Schwierigkeit der Implementierung wären im Vergleich leicht erhöht, das Verständnis bei den potenziell ungeschulten Betrachtern, vor allem Stakeholdern, gerade in kurzen Momenten einer Präsentation aber leicht gemindert. Baumdarstellungen anderer Anordnung, z.B. radialer, wären auch möglich, würden gut verstanden werden, zögen aber deutlich erhöhten Aufwand und Schwierigkeiten der Implementierung mit sich.

## 5.2 Anwendung Visualisierung Zwei

In dieser Visualisierung eines Parallelen Koordinaten Plots, kann der mittlere Manager in seiner Präsentation den oberen Managern und Stakeholder einen detaillierteren Blick auf die Verkaufszahlen der Videospiele eines Genres in den verschiedenen Regionen der Welt bieten. Der Blick der Betrachter wird wie zuvor beschrieben gelenkt. Zusätzlich finden sich wieder Erklärungen zu Einheiten und Darstellungsweisen auf der Webseite.

Der Präsentator wählt nun das zuvor festgelegte Genre *Action* aus und kann den Parallelen Koordinaten Plot mit den voreingestellten Achsenbelegungen und den gefilterten Videospielen als Datenpunkte betrachten. Schnell fällt der beinahezu allen Spielen sichtbare Abstieg der Verkaufszahlen in Japan auf. Um folgend Zusammenhänge zwischen den anderen Regionen besser



erkennen zu können, wird die Achsenbelegung wie im folgenden Bild gezeigt ausgewählt. Durch die Röntgenstrahlen nachempfunde Darstellungsweise lässt sich nun ein Muster erkennen. Zum einen befinden sich der Großteil der Videospiele im unteren Drittel der Verkaufszahlen in allen Regionen, was durch die Überlappungen der Datenpunkte und damit kräftigeren Grüntöne besser erkennbar ist. Einzig *Assassin's Creed: Unity* und vor allem *Grand Theft Auto V* von *Rockstar Games* stechen heraus. Wenn sich Spiele gut in Nordamerika verkauften, so verkauften sie sich in Europa in ähnlicher Weise, genauso wie im Rest der Welt und global. Dies wird auch durch unterschiedliche Achsenbelegungen bestätigt. Eine Korrelation über alle Regionen hinweg scheint gegeben. Einzig Japan bildet aus unbekannten Gründen eine Ausnahme, da die meisten Spiele hier wie anfangs erwähnt nicht so gut verkauft werden, wenn nicht sogar gar nicht. Gleichzeitig schneiden einige wenige Spiele, die in den anderen Regionen durchschnittlich verkauft werden, in Japan vergleichsweise gut ab.

Durch die Korrelation scheint eine Konzentration in diesem Genre auf eine bestimmte Region wenig Sinn zu ergeben, sodass eine breite Aufstellung und Nutzung der Beeinflussung nützlich ist. Gleichzeitig deutet die ansatzweise Bildung eines Clusters im unteren Drittel der Verkaufszahlen in allen Regionen auf starke Konkurrenz im Genre hin. Diese wird verstärkt durch die identifizierten Ausreißer, welche wie alle anderen Titel durch Hovern über den Datenpunkt gut ablesbar sind wie obiger Abbildung zu sehen. Ein neuer Titel sollte also sehr überzeugend sein oder eventuell doch in einem anderen Genre platziert werden. Sollte nach einer weiteren Bestätigung der Erkenntnisse im Scatterplot die Entscheidung zu weiteren Marktstudien in dem Genre getroffen werden, müsste anhand weiterer Einflussparameter unter anderem der Grund für die hohen Verkaufszahlen geprüft, sowie die Sättigung des globalen Marktes durch diesen Titel genauer analysiert werden.

Im Gegensatz zu *Action* gibt es im Genre *Sports* keine Ausreißer über alle Regionen hinweg und die Verkaufszahlen verteilen sich stärker über die Skala. Japan mit fast nur Verkaufszahlen von null bleibt bestehen. Eine Begründung kann die Visualisierung nicht liefern. Deutlich erkennbar ist jedoch, dass einige Spiele in Nordamerika besser verkauft werden als in Europa, im Rest der Welt und global aber wieder ähnlich, wenn auch etwas schlechter, als in Nordamerika. Gleichzeitig verkaufen sich Spiele, die in Europa gut verkauft werden, in Nordamerika etwas schlechter, im Rest der Welt und global aber wiederum besser. Daraus ergibt sich eine erste Tendenz, dass eine Konzentration auf eine der Regionen *North America* oder *Europe* sinnvoll ist, um wiederum die positiven Beeinflussungen dieser jeweiligen Regionen auf die restlichen Regionen zu nutzen. Sie wird im Scatterplot überprüft.

Der Vergleich mehrerer Dimensionen resp. mehrerer Attribute eines Videospieles wäre auch durch bspw. Icontechniken möglich. Diese Methode wäre jedoch schwerer umzusetzen und aufwändiger zu analysieren. Der Grund liegt in der Darstellung aller Dimensionen eines Videospieles auf einem Icon und einer entsprechend erschwerten Erkenntnis der Zusammenhänge und Muster über die Spiele hinweg. Auch Ausreißer sind mit dieser Technik nicht so gut erkennbar.

### 5.3 Anwendung Visualisierung Drei

Um zur dritten Visualisierung, dem Scatterplot, zu gelangen, wählt der mittlere Manager in seiner Präsentation den Plot im Drop-Down-Menü aus. Die zuvor getroffene Auswahl des Genres, zuletzt *Sports* bleibt bestehen und er wählt in einem weiteren Drop-Down-Menü die Belegung der Achsen auf *North America* und *Europe*.

Im Scatterplot werden die zuvor erkannten Cluster in den Verkaufszahlen im Vergleich von Nordamerika zu Europa deutlicher. Während bei einigen Spielen zwischen beiden Regionen ein deutlich positiv korreliertes Cluster sichtbar ist, scheinen die anderen Spiele ein Cluster ohne Korrelation zwischen den Regionen zu bilden. Dies bestätigt und verdeutlicht die Erkenntnisse aus der zweiten Visualisierung, dass sehr gute Verkaufszahlen von einigen Sportvideospiele in Nordamerika keinen Einfluss auf die Verkäufe in Europa haben. Die Titel im Cluster der positiv korrelierten Spiele scheinen sich jedoch gegenseitig positiv zu beeinflussen. Verweist der Präsentator mittels Hovern über die Datenpunkte auf die Titel der Videospiele, wird deutlich, dass diese Cluster mutmaßlich durch die unterschiedliche Popularität der Sportarten in beiden Regionen entsteht. So ist Football in Europa nicht so bekannt oder beliebt, Fußball aber in Nordamerika schon, wenn auch leicht weniger als in Europa.

Zuletzt interessieren sich die Teilnehmer der Präsentation für die Überprüfung der Erkenntnisse zum Genre *Action*, sodass die Filterung entsprechend angepasst wird. In allen möglichen Kombinationen der Achsenbelegung außer in denen mit Japan bestätigt sich die leicht bis vollständig positive Korrelation zwischen den jeweiligen Verkaufszahlen der Regionen. Weiterhin bildet sich in Nähe des Ursprungs im unteren Drittel der Skalen eine Andeutung eines Clusters, wie es auch im Parallelen Koordinaten Plot zu erkennen war. Im Scatterplot sind jedoch die einzelnen Datenpunkte etwas besser erkenn- und zur näheren Betrachtung auswählbar. Auch der zuvor erkannte Ausreißer sticht deutlich heraus. Zusätzlich bestätigt sich die nicht vorhandene Korrelation zwischen den jeweiligen Regionen und Japan. Die Titel des eigenen Publishers befinden sich immer im unteren Bereich des Clusters und scheinen vergleichsweise bisher nicht so gut verkauft worden zu sein. Sollte während der Präsentation der Wunsch nach einer gleichzeitigen Betrachtung des Baumdiagramms zur Übersicht der Publisher aufkommen, so kann mittels des Links am jeweiligen Seitenende in einem neuen Fenster dieser geöffnet werden und per Multitasking betrachtet werden.

Eine sinnvolle Alternative zu Scatterplots ist nicht gegeben, da sie die beste und am leichtesten zu verstehende Darstellung von Beziehungen zweier Attribute eines Datenpunktes ist.

Aufgrund der Erkenntnisse aller drei Visualisierungen können sich die oberen Manager und wichtigsten Stakeholder von *505 Games* nun dazu entscheiden, sich bei weiteren, (kosten-) intensiven Marktstudien mit detaillierteren Visualisierungen und weiteren Attributen auf das Genre *Sport* zu konzentrieren. Dort sind sie selbst noch nicht so stark präsent, es gibt weniger und in der Masse weniger starke Konkurrenz als bei *Action*. Weiterhin müssen nicht die Abhängigkeiten zwischen allen Regionen betrachtet werden, sodass eine Fokussierung und Abschöpfung eines Marktes Sinn ergibt. Eine Fokussierung auf den europäischen oder nordamerikanischen Markt ist

möglich und für ein kleineren Publisher wie *505 Games* sinnvoller. Durch die Ausrichtung auf den europäischen Markt kann *505 Games* jedoch die positive Korrelation mit dem nordamerikanischen Markt für seine Verkaufszahlen nutzen und damit auch ihn bedienen. Untersucht werden sollten künftig auch die Einflüsse für die vergleichsweise guten Verkaufszahlen für das eigene Spiel *Rocket League*, um über eine mögliche Fortsetzung zu entscheiden.

## 6 Verwandte Arbeiten

In diesem Kapitel wird eine knappe Literatursuche nach ähnlichen Anwendungen zu Videospielverkäufen oder Verkaufsdaten im Allgemeinen im Bereich der Informationsvisualisierung und Visual Analytics durchgeführt.

Der erste zu diskutierende Artikel ist *VizInteract: Rapid Data Exploration Through Multi-touch Interaction with Multi-dimensional Visualizations* von Chakraborty und Stuerzlinger.[1] In diesem Artikel wird *VizInteract* vorgestellt und getestet. Es dient der Multi-Touch Interaktion, um multidimensionale Datenvisualisierung schneller und einfacher konstruieren und mit ihr interagieren zu können. Es sind verschiedene Datenvisualisierungen implementiert, die mittels Verschieben und Übereinanderverschieben zu neuen Plots resultieren können. So können laut der Autoren z.B. zwei orthogonale Histogramme durch Übereinanderverschieben einen Scatterplot kreieren. Forschungsziel des konkreten Artikels die Beobachtung des Nutzerverhaltens der Anwendung einfacher touchbasierter Interaktionen.

Als Anwendungsfall wird für die Tester unter anderem ein Unterdatensatz der *Video Game Sales 2019* bereitgestellt, mit dem sie Visualisierungen mittels des Tools erstellen und damit vorgefertigte Analysefragen beantworten sollen. Der Datensatz ähnelt dem hier verwendeten, gleicht sich jedoch nicht. In *VizInteract* sind Histogramme, Scatterplots, Parallele Koordinaten Plots, Scatterplotmatrizen und Sterndiagramme möglich. Somit stimmen auch zwei der im Projekt verwendeten Visualisierungstechniken mit denen im Artikel überein.

Gemeinsamkeiten zwischen der Anwendung und Implementierung im vorliegenden Projekt und dem Artikel liegen in der Möglichkeit, Notwendigkeit und Umsetzung von Interaktionen für sinnvolle Visualisierungen, z.B. mittels Filtern. Weiterhin werden in beiden Fällen Scatterplots und Parallele Koordinaten eingesetzt und in den den Testern gestellten Aufgaben ähnlich angewandt wie im Anwendungsfall des Projektes. Durch den Umfang von *VizInteract* und die Tests im Artikel zeigt sich wiederum die Wichtigkeit, die die Autoren den Visual Analytics beimessen. Klar ersichtlich ist die Gemeinsamkeit der Nutzung der Videospielverkäufe mit einem sehr ähnlich aufgebauten Datensatz sowie Visualisierungen. Unterschiede ergeben sich aus den weiteren Visualisierungsmöglichkeiten des Tools, der Bedienbarkeit mit Touch-Oberflächen sowie der fortgeschrittenen Interaktion mittels Komposition der Techniken. Im Artikel wird das Tool anders als hier vorliegend mittels eines Anwendungsfalles auf seine Hauptmerkmale geprüft und seine Handhabung fokussiert.

Als zweites sei *Intelligent Visual Analytics Queries* von Hao et al. genannt.[3] In diesem Artikel

möchten die Autoren Analysten mittels ihrer *Intelligent Visual Analytics Query* unterstützen. Der Fokus liegt auf großen multidimensionalen Datensätzen, in die durch das Tool Einsicht in komplexe Muster, Phänomene und Ausreißer erlangt werden soll. Die angestrebte Anwendung beschreibt einen Analysten, der in einer noch unübersichtlichen Visualisierung der Daten einen Interessensbereich sowie die dazugehörigen Attribute auswählt. Anschließend sollen durch automatische analytische und visuelle Analysemethoden Charakteristiken und Beziehungen zu anderen Attributen und Datenpunkten identifiziert werden. Der im Artikel genutzte Anwendungsfall bezieht sich auf Verkaufszahlen von Produkten und Kundenkaufverhalten, da Verkaufsanalysten Produktverkäufe und Promotionen laut der Autoren korrelieren wollen.

Das vorgestellte Tool nutzt eine tabellenartig angelegte visuelle Karte, bei der in den Zeilen die Gruppen an Dimensionen und in den Spalten die Datenintervalle gespeichert werden. Die Farbe kennzeichnet den Attributwert für den Datenpunkt. So können Korrelationen und Ähnlichkeiten mehrerer Attribute erkannt werden. Weiterhin wird wie in der hier vorliegenden Arbeit der Parallele Koordinaten Plot visualisiert. Es können interessante Untergruppen von Daten ausgewählt und die paarweise Korrelation der gewählten Attribute berechnet werden. Die Achsen werden schließlich so angeordnet, dass hochkorrelierte Attribute nah beieinander sind. Auch Scatterplots werden genutzt, wobei diese hier so angeordnet werden, dass weiterhin multidimensionale statt zweidimensionale Daten abgebildet werden können.

Gemeinsamkeiten bestehen in der Auswahl der Visualisierungstechniken und der Anwendung dieser zum Erkennen von Mustern und Korrelationen in großen multidimensionalen Datensätzen zu Verkaufszahlen. Zusätzlich wird die Nützlichkeit dieser Techniken und der Visual Analytics an sich für die Auswertung und das Verständnis von Verkaufsdaten ähnlich groß eingeschätzt. Auch die Nutzung und Bewertung der Existenz von Interaktionsmöglichkeiten in den Visualisierungen durch Filter deckt sich mit denen der Autorin der vorliegenden Arbeit. Unterschiede befassen sich mit dem Einsatz von Scatterplots für mehr als zwei Dimensionen. Trotz der Ähnlichkeiten der Parallelen Koordinaten Plots ist die Anordnung der Achsen hier verändert und nimmt dem Nutzer Arbeit ab. So muss dieser nicht selbst die Achsen variabel aussuchen, um Muster und Korrelationen besser zu erkennen. Gleichzeitig fehlt ihm diese Flexibilität bei Hao et al. Wie schon im vorherigen Artikel wird weniger die Beantwortung einer Frage aus dem Anwendungsfall heraus behandelt, sondern das Tool allgemein vorgestellt.

## **7 Zusammenfassung und Ausblick**

In dieser Arbeit wurde mittels drei Visualisierungstechniken eine erste Marktanalyse der Verkaufszahlen des Videospielemarktes für die Konsole *XBoxOne* durchgeführt. Für die Zielgruppe des mittleren und oberen Managements als auch eingeschränkt der Stakeholder der Videospieleverlage, hier Publisher, wurden die vorliegenden Daten visuell aufbereitet. Durch die Implementierung eines expliziten Baumdiagramms konnte die gewünschte hierarchische Übersicht über den Markt sowie die Zusammenhänge zwischen Publisher, Genre und Titel umgesetzt werden.

Gleichzeitig konnte durch diese Technik die Konkurrenz betrachtet und abgeschätzt werden, welche Genres durch die quantitative Anzahl der Spiele und deren Publisher Potenzial bieten und im Folgenden näher betrachtet werden sollten. Mittels eines Parallelen Koordinaten Plots zur Darstellung mehrdimensionaler Daten konnten auch durch den Einsatz einer Art Röntgenstrahlentechnik Muster in den Verkaufszahlen von Videospielen eines Genres über mehrere Dimensionen, also Regionen, hinweg erkannt werden. Zuletzt konnten in einem klassischen Scatterplot die Erkenntnisse aus der vorherigen Visualisierung überprüft werden. Korrelationen zwischen zwei gewählten Regionen sowie Ausreißer und die Bildung von Clustern konnten durch diese Visualisierung detaillierter dargestellt und analysiert werden. Für ein verbessertes, komfortables und schnelles Anwendungserlebnis sowie Effektivität und Effizienz wurde die Auswahl der Genre mittels Drop-Down Menü interaktiv gestaltet sowie beim komfortablen Umschalten der Visualisierungen ebenso mittels Drop-Down Menü erhalten. Aus demselben Grund wurde auch ein Beibehalten der flexiblen und individuellen Achsenauswahl für den Parallelen Koordinaten Plot sowie den Scatterplot integriert.

Der Mehrwert der Visualisierungen liegt für die Zielgruppe vor allem im schnellen Verständnis und Überblick über den aktuellen Videospiegelmarkt mit Fokus auf Verkaufszahlen für die *XBoxOne*. Zusammenhänge und Muster zwischen den Regionen in den Genres können ohne Vorwissen im Bereich der Visual Analytics abgelesen, analysiert und präsentiert werden. So bieten die Visualisierungen erste Erkenntnisse für die Zielgruppe bezüglich neuer Möglichkeiten und Chancen im Markt sowie Ansatzpunkte für grobe Strategieentscheidungen. Sie dienen als Entscheidungsunterstützung für die Beauftragung detaillierterer, fokussierter und teurerer Marktstudien für neue Investitionen.

Sinnvolle Erweiterungen der Daten wären der Einbezug von (Nutzer-)Kritiken und deren mögliche Auswirkungen und Korrelationen auf die Verkaufszahlen sowie zwischen den Regionen. Auch die Analyse bspw. auf Korrelationen nicht öffentlich zugänglicher Daten zu Entwicklungskosten und Umsätzen pro Spiel bieten Potenzial. Dies ist jedoch Teil einer angesprochenen weiterführenden, detaillierten Marktstudie. Interessant wäre eine weitere Unterteilung der Regionen, vor allem der unter *Rest of World* zusammengefassten Regionen wie Afrika, Asien und Südamerika. Bezüglich der Visualisierungstechniken kann eine Interaktion mit Klick auf einen Genreknoten im Baumdiagramm und die Anzeige des dazu passenden Parallelen Koordinaten- und Scatterplots Sinn ergeben. Deutlich mehr Potenzial bieten jedoch Ansätze aus den *Verwandten Arbeiten*. So ist ein automatisches Anordnen der Achsen je nach Korrelationsstärke im Parallelen Koordinaten Plots wie bei Hao et al. interessant.[3] Zur Verbesserung der Interaktionen gerade auf der heutzutage vielfältigen Auswahl von Endgeräten ergibt eine Implementierung einer Komposition von Visualisierungen durch Zusammenschieben wie bei Chakraborty und Stuerzlinger Sinn.[1]

Insgesamt konnten die durch die mögliche Anwendung entstandenen Anforderungen an die implementierten Visualisierungen gut umgesetzt werden.

## Anhang: Git-Historie

### Literatur

- [1] Supratim Chakraborty und Wolfgang Stuerzlinger. “VizInteract: Rapid Data Exploration Through Multi-touch Interaction with Multi-dimensional Visualizations”. In: *Human-Computer Interaction - INTERACT 2021*. Hrsg. von Carmelo Ardito u. a. LNCS sublibrary, SL 3, Information systems and applications, incl. internet/web, and HCI. Cham, Switzerland: Springer, 2021, S. 610–632. ISBN: 978-3-030-85613-7.
- [2] Jürgen Fleig. *Marktanalyse und Marktforschung: Definition, Zweck und Beispiele*. 2020. URL: <https://www.business-wissen.de/hb/marktanalyse-und-marktforschung-definition-zweck-und-beispiele/> (besucht am 25.11.2022).
- [3] Ming C. Hao u. a. “Intelligent Visual Analytics Queries”. In: *IEEE Symposium on Visual Analytics Science and Technology, 2007*. Hrsg. von William Ribarsky. Piscataway, NJ: IEEE Service Center, 2007, S. 91–98. ISBN: 978-1-4244-1659-2.
- [4] o.A. *Videospiele: Weltweit*. 2022. URL: <https://de.statista.com/outlook/dmo/digitale-medien/videospiele/weltweit#umsatz> (besucht am 25.11.2022).
- [5] o.A. *XBox Series X/S Jubiläum: Die Bilanz nach dem ersten Jahr*. 2021. URL: <https://www.gameswirtschaft.de/wirtschaft/xbox-series-x-s-geburtstag-analyse/> (besucht am 25.11.2022).
- [6] SID\_TWR. *Video Games Sales Dataset: Video Games Sales & Game Ratings Data Scraped from VzCharts*. URL: [https://www.kaggle.com/datasets/sidtwr/videogames-sales-dataset?select=XboxOne\\_GameSales.csv](https://www.kaggle.com/datasets/sidtwr/videogames-sales-dataset?select=XboxOne_GameSales.csv) (besucht am 13.11.2022).
- [7] F. Tenzer. *Monatliche Verkaufszahlen der Xbox One in Europa bis September 2022*. 2022. URL: <https://de.statista.com/statistik/daten/studie/311733/umfrage/absatz-der-xbox-one-pro-monat-in-europa/> (besucht am 25.11.2022).