

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

## **Marktanalyse des Videospielmарktes**

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

Lena Arloth

Matrikelnummer: 217207784

7. Dezember 2022

---

Link zum GitHub-Repository:

<https://github.com/Lena-Ar/Info-Vis>

Letzter Commit:

()

Link zur öffentlich zugänglichen Website (GitHub-Pages):

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Anwendungshintergrund . . . . .	1
1.2	Zielgruppe . . . . .	2
1.3	Überblick und Beiträge . . . . .	3
<b>2</b>	<b>Daten</b>	<b>3</b>
2.1	Bereitstellung und Vorverarbeitung der Daten . . . . .	4
<b>3</b>	<b>Visualisierungen</b>	<b>5</b>
3.1	Analyse der Anwendungsaufgaben . . . . .	5
3.2	Anforderungen an die Visualisierungen . . . . .	7
3.3	Präsentation der Visualisierungen . . . . .	8
3.3.1	Visualisierung Eins . . . . .	8
3.3.2	Visualisierung Zwei . . . . .	9
3.3.3	Visualisierung Drei . . . . .	11
3.4	Interaktion . . . . .	12
<b>4</b>	<b>Implementierung</b>	<b>13</b>
4.1	Data.elm . . . . .	13
4.2	TreeHierarchy.elm . . . . .	14
4.3	ParallelPlot.elm . . . . .	15
4.4	Scatterplot.elm . . . . .	16
4.5	MainScatterParallel.elm . . . . .	17
<b>5</b>	<b>Anwendungsfälle</b>	<b>20</b>
5.1	Anwendung Visualisierung Eins . . . . .	20
5.2	Anwendung Visualisierung Zwei . . . . .	21
5.3	Anwendung Visualisierung Drei . . . . .	22
<b>6</b>	<b>Verwandte Arbeiten</b>	<b>23</b>
<b>7</b>	<b>Zusammenfassung und Ausblick</b>	<b>25</b>

# 1 Einleitung

Der weltweite Videospielmarkt 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.[10] Entsprechend viel Geld und Potenzial durch Wachstum verbirgt sich in ihm. Gleichzeitig benötigt es für die Entwicklung neuer Videospiele hohe Investitionen vonseiten 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 Videospieltitel aufgeschlüsselt nach Genre weltweit.

Als Problemstellung ergibt sich das fehlende Wissen der Entscheider innerhalb der Publisher, in welche Videospieltitel in welchem Genre und mit welcher Regionsausrichtung sie künftig investieren sollten, um das Potenzial des großen und wachsenden Videospielmarktes 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 einer Plattform in der Videospielindustrie 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.[4] 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.

Es 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 Videospieltitel 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.[4] Dabei kann sie detaillierter oder oberflächlicher sein. In dieser Anwendung ist eine oberflächliche Analyse zur Ersparnis von Zeit und Geld geplant, mittels derer monetär interessante Genres und Regionen identifiziert werden können. Aufgründessen kann eine Investition in eine detaillierte, deutlich aufwenigere und damit teure Marktstudie in den zuvor identifizierten Genres und mitunter Regionen in Auftrag gegeben werden. Diese kann zu einer deutlich fundierteren Bestätigung oder Ablehnung der Vorabentscheidung beitragen.

Der Videospielmarkt 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.[15] 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.[11] Somit behalten die Verkaufszahlen der *XBoxOne* ihre wenn auch leicht verminderte Aussagekraft. Der Sinn einer übersichtsartigen, 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. Erstens wird das Bedürfnis nach übersichtsartigen Informationen zur Videospieleindustrie im Sinne der eigenen Position und der Konkurrenz für die *XBoxOne* bezüglich der Verbindungen von Publisher, Genre und Videospiele adressiert. Zweitens werden Informationsbedürfnisse nach Zusammenhängen und Mustern zwischen Regionen in den jeweiligen Genres befriedigt, um daraus strategische Entscheidungen ableiten zu können. Drittens 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.[14] Die Visualisierungen entwickeln sich vom Groben zum Detaillierten. In einer ersten Visualisierung wird zur Unterstützung einer Übersicht über die Publisher, Genre und Videospieletitel 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*. [14] 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 *Metacrtitics* sind hier nicht enthalten. Der Originaldatensatz liegt als CSV-Datei vor und beinhaltet zehn Spalten mit 613 einzelnen Positionen.

Zu jeder *Position* sind der Videospieldname 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 Ablauf 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 Videospieles 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 aufgrund dessen fehlerhaft und haben in der Visualisierung keine Verwendung. Zum Ausschluss 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 übersichtsartige 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

Nachfolgend wird eine Analyse der Anwendungsaufgaben zur Lösung des Zielproblems durchgeführt, Anforderungen an die Visualisierungen abgeleitet und diese präsentiert.

#### 3.1 Analyse der Anwendungsaufgaben

Das Zielproblem der konkreten Anwendungsaufgaben ist eine grobgranulare Marktanalyse des Videospielesmarktes der *XBoxOne* als Entscheidungsunterstützung zur Investition in detaillierte Studien und folglich neue Videospiele. Fokussiert wird die Frage, welche Genres und möglicher-

weise Spezifizierungen auf bestimmte Regionen dort untersucht werden sollen. Zur Problemlösung sollen die Verlagshäuser sowie ihre Konkurrenz betrachtet werden als auch Verkaufschancen, Zusammenhänge und Auffälligkeiten in den einzelnen Genres zwischen den Regionen der Welt. Mit Rücksicht auf die Zielgruppe ist eine Lösung im Kontext von Präsentationen sowie Kosten- und Zeitminimierung wichtig. Weiterhin ist die sprachliche Diversität der Zielgruppe durch die globale Verteilung der Publisher zu berücksichtigen.

Als erste Anwendungsaufgabe ergibt sich die Schaffung eines Marktüberblicks als Ansatzpunkt weiterer Entscheidungen und erste Eingrenzungen der zu untersuchenden Genres. Das Management der Publisher benötigt eine Übersicht über die eigenen und konkurrierende Angebote, die insbesondere für größere, breit aufgestellte Verleger sinnvoll ist. In dieser Übersicht sollten die Publisher, ihre abgedeckten Genres und die in ihnen kategorisierten Videospieltitel ersichtlich sein. Je nach Strategie des betrachtenden Publishers können bspw. aufgrund der geringen Anzahl der eigenen Spiele in einem Genre oder aber der hohen Anzahl der konkurrierenden Titel in einem solchen Vorentscheidungen zur Eingrenzung des mittels nachfolgender Anwendungsaufgabe näher zu untersuchenden Genres getroffen werden. Zur Lösung der Anwendungsaufgabe soll ein mentales Modell von Beziehungen und Hierarchien zwischen Publishern, Genre und Titel bildlich entstehen, bestenfalls auf einen Blick. Dies könnte auch schriftlich durch Tabellen oder Texte umgangen werden, jedoch unterstützt das mentale Modell das Verständnis und beschleunigt den Vorentscheid ungemein. Besonders im Rahmen einer Präsentation ist es unumgänglich.

Als zweite Anwendungsaufgabe ergibt sich die Entscheidung für eine Investition in ein Genre mit Bezug auf die Regionen mittels einer Analyse von Verkaufschancen, Zusammenhänge und Auffälligkeiten in diesem. Es ist wichtig, erste Auffälligkeiten in allen Regionen der Welt zu erkennen, um ein Genre für eine Investition in eine folgende Marktstudie zu fokussieren. Dazu zählen Spiele, die auffällig gut verkauft wurden und eine starke Konkurrenz darstellen sowie Muster in den Verkaufszahlen, die aus mehreren Videospielen im Genre in den Regionen der Welt gebildet werden wie bspw. viele Spiele, die in allen Regionen außer einer mittelmäßig gut verkauft wurden. Weiterhin ist es für eine mögliche Spezifikation der künftigen Marktstudie auf eine Region sinnvoll, diese Muster genauer durch Untersuchung nur besonders interessanter Regionen zu prüfen. Somit können Abhängigkeiten besonders zwischen zwei Regionen detailliert analysiert werden. Benötigt wird zudem eine Erkennbarkeit der konkreten Informationen der Spieletitel. Bei der Bearbeitung können zwei mentale Modelle helfen. Zum einen soll mental eine Art Sicht von oben auf die Welt mit allen Regionen entstehen, die durch die Videospiele als Datenpunkte verbunden sind. Durch Häufungen von Videospielen und Ähnlichkeiten in den Verbindungen vieler Spiele, sollen mental Muster und Cluster über alle Regionen entstehen. Zum anderen ist ein mentales Modell mit erhöhtem Detaillierungsgrad zur Lösung der Anwendungsaufgabe sinnvoll, mit dem je zwei zuvor als interessant erkannte Regionen vergrößert werden. Hierbei sollen durch Punkte in Raum die Verkaufszahlen mental positioniert und damit vergleichbar werden sowie Korrelationen in ihrer Art leichter bestimmbar sein. Durch Punkthaufen und deutlich anders positionierte Punkte sollen auch die vorher erkannten Muster, Cluster und Ausreißer im Detail



prüfbar sein. Visualisierungen zum Aufbau dieser mentalen Modelle sind elementar, da ohne sie teils Analysen der gewünschten Aspekte nur sehr schwer möglich sind, vor allem die Erkennung von Clustern und Korrelationen.

Wichtig zur Lösung des Zielproblems ist das intuitive, schnelle und eindeutig bei allen Managern der Publisher Entstehen der mentalen Modelle zur einheitlichen Analyse der Daten gerade bei Präsentationen oder dem kurzfristigen Gebrauch der Visualisierungen. Zu beachten ist auch, dass schnelle Entscheidungen Wettbewerbsvorteile bedeuten können. Die Analyse der Anwendungsaufgaben ergibt zudem möglichst geringe monetäre und zeitliche Aufwände zur Erstellung der Visualisierungen zur Unterstützung der mentalen Modelle. Die Begründung liegt in der gewünschten Kostenersparnis durch eine vorab günstigere, grobgranulare Marktanalyse zur Investitionsentscheidung in detaillierte, teure und zeitintensive Marktanalysen höheren Umfangs.

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

Über alle Anwendungsaufgaben hinweg stellen sich die Anforderungen nach Expressivität, Effektivität und Angemessenheit nach Schuman und Müller.[13, 9ff.] Für die gewünschte Expressivität ist das Verständnis der Visualisierungen ohne Missverständnisse oder Doppel- bzw. Mehrdeutigkeiten der Daten nötig. Es sollen nur die in den Daten enthaltenen Aussagen visualisiert werden. Es werden drei visuelle Ebenen der Informationen definiert, von denen hier die obere Ebene zutrifft.[1] Für eine gute Expressivität im Rahmen dieser oberen Ebene ist entsprechend die Visualisierung expliziter Fakten zu den Videospielen und Verkaufszahlen sowie versteckte Zusammenhänge. Wie zuvor analysiert sind hier vor allem Offenlegung von Hierarchien und Mustern, Abhängigkeiten und Ausreißern zwischen Regionen in einem Genre zu nennen. Zur Erfüllung der Effektivität ist ein intuitives, schnelles Verständnis der Visualisierungen gefordert. Die Wichtigkeit dieser Anforderung begründet sich im zuvor analysierten Zeitdruck durch die Anwendung in Präsentationen sowie der Entscheidungen. Aufgrund der mit wenig spezifischem Vorwissen im Bereich Visual Analytics ausgestatteten Zielgruppe bedeutet Effektivität auch eine geringere Komplexität der Darstellungen, ergo intuitiv für die Zielgruppe verständliche Visualisierungen vor allem von Korrelationen und Mustern. Durch die der Zielgruppe zugrunde liegende minimal verfügbare Zeit zum Verständnis liegt der Fokus klar auf zeitlicher Effizienz bei bestmöglicher Effektivität. Durch die Anwendung der Visualisierungen zur Entscheidungsunterstützung des Fokus weiterer Marktstudien ist besonders die Anforderung an Angemessenheit zentral. Die Visualisierungen müssen wenig Zeit- und Kostenaufwändig sein, um einen Vorteil in dem vorgestellten Verfahren zu bieten. Dazu ist es notwendig, den Umfang und den Aufwand der Visualisierungen zum Erreichen der Expressivität für die Darstellung der oberen visuellen Ebene der Informationen auf niedrigerem Niveau zu halten.

Bezüglich der Farbgebung der Visualisierungen und Website ist eine Rücksichtnahme auf mögliche Sehschwächen der Zielgruppe nötig. Ohne Expressivität und Effektivität negativ zu beeinflussen, ist eine Beschränkung auf eine Farbgruppe sinnvoll. Untergeordnet sind außerdem Anforderungen an Ästhetik und indirekte Leitung des Blicks über die Website durch gerade

erkennbare Abstufungen in Farbe und Kontrast, sogenannte *Just Noticable Differences*. Zum einheitlichen Verständnis der Visualisierungen die Diversität der Zielgruppe berücksichtigend, wird die Sprache Englisch festgelegt.

Die Analyse der zweiten Anwendungsaufgabe ergab zuvor die Notwendigkeit der Auswahl möglicher Genres zur Entscheidung für Investitionen in Marktstudien in bestimmten Genres. Folglich gilt die Anforderung an einen Filter zur Auswahl der Genres in den zwei Visualisierungen der Verkaufszahlen. Aus der Analyse jener zweiten Anwendungsaufgabe folgt außerdem die Anforderung an die dritte Visualisierung mit der Auswahl von zwei Regionen zum Vergleich.

Konkret muss die erste Visualisierung als Ausgangspunkt der Anwendung die in der Analyse der ersten Anwendungsaufgabe geforderten Beziehungen zwischen Publishern, Genres und Titel abbilden und auf einen Blick für die Zielgruppe sichtbar machen. An die zweite Visualisierung stellt sich die Anforderung nach einer ganzheitlichen Sicht über alle Attribute der Regionen sowie die Aufdeckung von in den Verkaufszahlen der Regionen versteckten Mustern, Clusterbildungen und möglichen Ausreißern über alle Regionen hinweg. Die Abbildung mehrdimensionaler Daten ist essentiell. Zur gewünschten Vergrößerung der vorherigen Ansicht auf zwei interessante Attribute, ist für die dritte Visualisierung die Ansicht zweidimensionaler Daten ausreichend. Jedoch müssen die Attribute resp. Regionen dazu variabel wählbar sein. An die dritte Visualisierung wird vor allem die Anforderung nach detaillierter Darstellung von Korrelationen und deren Arten gestellt. Nützlich ist eine Überprüfung der in der zweiten Visualisierung sichtbaren Muster und Cluster in den zwei gewählten Dimensionen.

### **3.3 Präsentation der Visualisierungen**

Es werden drei Visualisierungstechniken zur Erfüllung der Anforderungen gewählt und präsentiert. Zur Datenpunktmarkierung und -hervorhebung sowie der Leitung des Blicks über die Website werden neben Schwarz-Weiß werden zwei harmonische Grüntöne in unterschiedlichen Konstraten und Opazitäten genutzt.

#### **3.3.1 Visualisierung Eins**

Zur Erfüllung der Anforderungen an die Visualisierung der ersten Anwendungsaufgabe wird ein explizites Baumdiagramm gewählt. Mittels dieser können hierarchische Beziehungen zwischen Attributwerten als Knoten in Form eines Baumes mit einem Wurzelknoten abgebildet werden. Hierarchische Beziehungen zwischen den Knoten seien durch gerade Linien, Kurven oder Bögen gekennzeichnet, die Knoten miteinander verbinden.[12, S. 1]

Pfade zwischen den Knoten, sowie die Positionierung von Eltern über Kindknoten stellen hier die hierarchische Beziehung dar. Publisher sind Elternknoten der Genres, die wiederum Elternknoten der Videospieltitel sind. Zur Verhinderung von ungewollten Kreuzungen der Pfade zwischen Publishern und Genres, wird letzteren zur eindeutigen Identifikation das Kürzel des zugehörigen Publishers hinzugefügt. Andernfalls verlöre die Visualisierung ihren Zweck der

Übersicht und einfachen Betrachtung der eigenen abgedeckten Genres und in ihnen verordnete Titel sowie der Vergleich mit der Konkurrenz. Eine erste Eingrenzung der Genres durch oben genannte Analysen ist so möglich.

Die aufgeführten Anforderungen an Expressivität, Effektivität und Angemessenheit werden durch ein explizites Baumdiagramm gut erfüllt. So ist eine zweifelsfreie Darstellung der Beziehungen und Abbildung der Hierarchien möglich, ohne unpassende Assoziationen zu wecken. Alle Informationen, die hier für die obere visuelle Ebene der Informationen nötig sind, werden übermittelt. In den Daten versteckte Beziehungen werden offengelegt und konkrete Details durch die Anzeige der Knotenbezeichnungen bei Hovern mit der Maus aufgeführt. Durch das intuitive, schnelle Verständnis ohne notwendige Erklärungen und Legenden, die überblicksartige Ansicht und die weite Verbreitung solcher Diagramme in vielen unterschiedlichen Bereichen ist auch die Effektivität der Visualisierung hinreichend gegeben. Die durch Hovern angezeigten expliziten Informationen sowie die schnelle quantitative Übersicht über Angebote durch Zählung der von einem Elternknoten ausgehenden Pfade trägt unter anderem dazu bei. Weiterhin ist diese Visualisierungstechnik durch den eher als gering einzuschätzenden monetären und zeitlichen Aufwand zur Erstellung bei starker Expressivität und Effektivität angemessen.

Wie erwähnt, wird durch das explizite Baumdiagramm die Anforderung an die Darstellung von hierarchischen Beziehungen erfüllt. Durch den Einsatz des Walker-Algorithmus wird die Effektivität dank einheitlicher Positionierung der Knoten nochmals verbessert.[16, 688ff.] Vermindert wird sie gleichzeitig durch die hohe Anzahl an Knoten, die zu einem breiteren, leicht weniger übersichtlichen Diagramm führen. Durch das Duplizieren des Diagramms mit veränderter Größe wird der Nachteil der größeren Menge an Knoten reduziert.

Hyperbolische Bäume sind eine Alternative zum expliziten Baumdiagramm. Durch die Zeichnung der Beziehungen und Knoten im hyperbolsichen Raum und einen wachsenden Kreisumfang bietet diese Technik mehr Platz für größere Hierarchien.[8, 2f.] Jedoch zöge diese Alternative einen höheren (Kosten-)Aufwand mit sich, weniger Intuitivität in der Erkennung der eigenen und konkurrierenden Angebote sowie eine suboptimale quantitative Übersicht der qualitativen Daten. Implizite Baumdiagramme sind eine weitere Alternative, bei der die Eltern-Kind-Beziehung durch Umschließung des Kindknotens durch den Elternknoten visualisiert werde.[12, S. 394] Sie sind platzsparender, aber durch einen geringeren Bekanntheitsgrad in der Zielgruppe weniger intuitiv verstehbar. Als Überblick bieten sie zudem weniger schnell eine Übersicht über eigene und konkurrierende Angebote statt eines Entlangfahrens der Kanten mit dem Auge.

### **3.3.2 Visualisierung Zwei**

Für die erste Teilbearbeitung der zweiten Anwendungsaufgabe wird der Parallele Koordinaten Plot gewählt, da er sich gut für Visualisierungen von Datenverteilungen und Abhängigkeiten zwischen Attributen eigne.[3, S. 11] In positiver Richtung der Y-Achse werden parallel zueinander die in ihrem Wertebereich linear skalierten Achsen positioniert. Ein Datenpunkt hat verschiedene Koordinaten, mittels derer Polygonzüge konstruiert werden. Die Schnittpunkte mit den

jeweiligen Achse zeigen den Attributwert.[7, 25f.] [3, S. 11]

Auf den parallelen Achsen werden die Wertebereiche der Verkaufszahlen der fünf verschiedenen Regionen skaliert und nebeneinander angeordnet abgebildet. Die Schnittpunkte der Polygonzüge, resp. Datenpunkte, mit den Achsen kodieren die jeweiligen Attributausprägungen in den Regionen. Mehrdimensionalität und ein ganzheitlicher Blick werden erfüllt.

Es können sowohl die expliziten Details zu Videospielen und ihren Verkaufszahlen als auch die in den Daten versteckten Informationen zu Muster- und Clusterbildung, Abhängigkeiten zwischen mehreren Regionen und Ausreißer offengelegt werden. Ersteres wird durch die Anzeige der Informationen bei Hovern über ein Polygonzug über dem Plot visualisiert. Zweiteres durch die Verbindung der Attributausprägungen eines Titels durch Polygonzüge sowie die sich durch Überschneidungen, Überlappungen und Ansammlungen der Linien bildenden Muster. Weiterhin eignet sich die Technik durch die Abbildung der Attributwerte auf Position und in gewissem Maße Orientierung gut für quantitative Daten.[1] Durch eine freie Wahl der Achsenbelegungen mittels Drop-Down-Menü aller Achsen können zudem weitere interessante Muster, um den gesamten erwünschten Blick von oben zu erhalten, erkannt werden. Das Problem der laut Lehman bestehenden Darstellung von Abhängigkeiten nur zwischen drei benachbarten Achsen kann somit umgangen werden.[9, S. 594] Die Expressivität wird insgesamt erfüllt.

Muster, Cluster und Ausreißer lassen sich grundlegend gut durch die Polygonzüge erkennen. Nichtsdestotrotz wird die Effektivität leicht gemindert, da das korrekte Verständnis der Kodierung der Daten einigen Personen der Zielgruppe schwieriger fallen kann. Umso wichtiger ist die zuvor beschriebene Möglichkeit zur Wahl der Achsenbelegung sowie prägnanter Erklärtexte. Bei großen Datensätzen ist es in Parallelen Koordinaten Plots problematisch, einzelne Polygonzüge durch Überlagerungen zu verfolgen. Durch ein Hervorheben der Linie und der Informationsanzeige beim Hovern kann dies gelöst werden. Gleichzeitig erhalten alle Polygonzüge eine grüne Färbung mit geringerer Opazität, sodass Schnittstellen und Häufungen von Liniensegmenten durch einen kräftigeren Ton leicht erkennbar sind. Diese Art von Röntgenstrahlentechnik kann die Erkennung von Mustern und Clustern verbessern und verbessert die Effektivität wieder. Das Design der Visualisierung mittels Filterung und freier Wahl der Achsenbelegung verhindert zwar nicht die vielen Kombinationsmöglichkeiten, jedoch die dabei mittels einfachem Achsentauch auftretende Überforderung. Parallele Koordinaten Plots sind zwar etwas (kosten-)aufwändiger zu erstellen, bleiben aber trotzdem im Kontext der Anwendungsaufgabe angemessen.

Eine Alternativen zu Parallelen Koordinaten Plots stellen Icon-Techniken, bspw. Star Plots oder Stickfigures dar. In Star Plots seien die Dimensionen als eckige Achsen ausgehend von einem Kreismittelpunkt mit einer äußeren Verbindungslinie zur Repräsentation der Attributausprägung jeder der Dimensionen visualisiert. Dabei steigt jedoch die Anzahl der zu erstellenden Glyphen mit steigender Menge an Datenpunkten stark, da sie nur je einen Datenpunkt repräsentieren können.[3, S. 20] Ein ähnliches Problem existiert auch bei Stickfigures, die zusätzlich durch zwei Attribute in einem Koordinatensystem mit weiteren positioniert werden. Zwar können so grobe Muster und Cluster erkannt werden, die anderen Anforderungen bspw. kon-

krete Informationen zu Datenpunkten werden allerdings unzureichend erfüllt. Intuitivität und Effektivität sind im vorherrschenden Kontext unzureichend, womit Icon-Techniken alternativ ungeeignet sind.

### 3.3.3 Visualisierung Drei

Zur vollständigen Beantwortung der zweiten Anwendungsaufgabe wird eine weitere Visualisierung benötigt, für die die Technik des Scatterplot gewählt wird. In einem kartesischen Koordinatensystem mit orthogonaler X- und Y-Achse werden X- und Y-Werte zweier variabler Attribute eines Datenpunkts zu bivariaten Paaren zusammengefasst, die als Punkte in diesem abgetragen werden. Durch die Position werden die Attributwerte visualisiert.[5, S. 103] So werden, auch durch Abbildung vieler Datenpunkte, die Form von Abhängigkeiten, Muster und Ausreißer zwischen zwei Attributen sichtbar.[3, S. 9] [5, S. 103] Die Schwäche des Scatterplots durch Verlust an Informationen von mehreren Dimensionen wird in diesem Fall durch eine Detaillierung und Überprüfung der Erkenntnisse des Parallelen Koordinaten Plots im Einzelnen zum Vorteil.[3, S. 9] [17, S. 93]

Im Design dieses Scatterplots ist die Belegung der Achsen zur Darstellung der Wertebereiche der Attribute *North America*, *Europe*, *Japan*, *Rest of World* und *Global* frei wählbar, um den Anforderungen nach Auswahl der interessanten Attribute des Parallelen Koordinaten Plots zur Überprüfung und Detaillierung dieses nachzukommen. Weiterhin sind somit alle möglichen Abhängigkeiten zwischen Attributkombinationen untersuchbar. Die Datenpunkte werden zur visuellen Ablesbarkeit wie oben beschrieben als Attributwert-Kombinationen kodiert, an den Achsen skaliert und als klassische Kreise im Raum des Koordinatensystems positioniert. Quantitative Daten lassen sich am besten auf Position und untergeordnet auch auf Orientierung abbilden, was der Scatterplot bestens umsetzt.[1]

Die Anforderungen an die Expressivität in Bezug auf die obere visuelle Ebene der Informationen werden durch die unmissverständliche Zuweisung der Videospieldaten zum Punkt im Diagramm mittels Hovern sowie die Visualisierung von versteckten Informationen erfüllt. Letztere bestehen aus dem deutlichen Aufzeigen von Korrelationen und ihrer Art, Clustern und Ausreißern zwischen zwei wählbaren Attributen. Weiterhin lässt sich mit ihm die gewünschte Vergrößerung der multidimensionalen Daten auf eine Ansicht von nur zwei Attributen und dem damit verbundenen höheren Detaillierungsgrad umsetzen. So kann die Analyse des komplizierteren Parallelen Koordinaten Plots überprüft und Fehler erkannt werden. Detailfragen zur finalen Entscheidung für ein weiter zu fokussierendes Genre oder eventuelle Korrelationen zwischen Regionen in diesem, die hinder- oder förderlich sein können, werden beantwortet. Durch eine deutliche räumliche Abgrenzung von Punkthaufen jeder Form können Cluster erkannt und bestätigt sowie einzelne Ausreißer bestimmt werden. Diese konkreten Anforderungen werden also auch erfüllt.

Auch Effektivität und Angemessenheit sind klar gegeben. Ersteres ergibt sich aus der starken Verbreitung der Technik in allen Bereichen und der unmissverständlichen, sich durch die Posi-

tionierung der Punkte ergebenden Darstellung. Damit visualisiert er intuitiv und schnell lesbar eine Fülle an detaillierten statistischen Informationen. Der Scatterplot ist durch seine Trivialität in der Abbildung zweier quantitativer Attribute angemessen. Dadurch ist er vergleichsweise unkompliziert und kostengünstig bei hoher Expressivität und Effektivität.

In diesem Kontext sind keine sinnvollen Alternativen möglich. Korrelationstabellen listen die Abhängigkeiten nur tabellarisch auf, visualisieren sie also nicht mittels einer spezifischen Technik. Mittels QQ-Plots können zwei statistische Variablen miteinander verglichen werden, jedoch mittels der geordneten Gegenüberstellung ihrer Quantile zur Erkennung von Verteilungen. Diese dienen nicht der Lösung des Zielproblem, könnten aber nach der Entscheidung für eine Investition in detaillierte Marktstudien in einem Genre dort Verwendung finden. So sei der Scatterplot "the most versatile, polymorphic, and generally useful"[5, S. 103] Visualisierung für statistische Daten und eignet sich damit perfekt zur abschließenden Beantwortung der zweiten Anwendungsaufgabe und des gesamten Zielproblems.

### 3.4 Interaktion

In jeder der Visualisierungen sind Interaktionen mittels Hovern über die Datenpunkte zur Anzeige konkreter Informationen im jeweiligen Kontext des Plots sowie zum Aufleuchten des Datenpunkts zur besseren Erkennung möglich. Dies dient der Erfüllung der Anforderungen an stets mögliche Detailansichten sowie der Verbesserung der Effektivität und Effizienz wie in den vorherigen Kapiteln beschrieben.

Auch in jeder einzelner Visualisierung sind Interaktionsmöglichkeiten integriert. So kann im expliziten Baumdiagramm der vergrößerte Baum mittels Scrollbar verschoben werden und mithilfe eines Links zu den detaillierteren Visualisierungen gewechselt werden. Von jeder der detaillierten Plots kann der Anwender durch einen Link am unteren Ende der Website schnell einen weiteren Tab zur Rückkehr zum Baumdiagramm als Ausgangspunkt und Übersicht für bspw. einen Abgleich öffnen. Im Parallelen Koordinaten Plot sind die Achsen frei belegbar dank eines Drop-Down-Menüs für jede der Achsen. Wie in Kapitel 3.3.2 beschrieben, können so Überforderung, Unübersichtlichkeit und Zeitverlust durch bloßen Achsentausch verhindert werden sowie die Abhängigkeiten zwischen den Regionen in jeder Kombination untersucht und versteckte Details entdeckt werden. Die Betrachtung der Abhängigkeiten zwischen nur drei benachbarten Dimensionen ist damit weniger problematisch. Auch der Scatterplot bietet die Interaktion mit je einem Button als Drop-Down-Menü zur freien Achsenbelegung, um die detaillierte Untersuchung von Abhängigkeiten und den Vergleich zwischen zwei beliebigen Attributen zu ermöglichen. Dies dient auch der gewünschten Vergrößerung und Detaillierung der vorherigen Visualisierung.

Ist der Anwender bei den detaillierten Plots angelangt, kann er mit einem Button als Drop-Down-Menü interagieren und so leicht und schnell zwischen allen Visualisierungen wechseln und die entsprechende wird angezeigt. Bei einem Wechsel zum Baumdiagramm leitet der Link zur eigentlichen Visualisierung dieses weiter. Für eine verbesserte Effektivität, Zeitersparnis und Komfort bleiben die Achsenbelegungen bei einem Plotwechsel mittels Drop-Down-Menü

gespeichert und müssen bei erneutem Wechsel zurück nicht wieder neu eingestellt werden. Welche Attribute auf den Achsen ausgewählt sind, ist durch eine sich anpassende Unterüberschrift des Plots gekennzeichnet.

Zuletzt wird die Anforderung an einen Filter zur Auswahl der Genres und entsprechender Eingrenzung der Stichprobe in den Visualisierungen Paralleler Koordinaten Plot und Scatterplot umgesetzt. Hierzu gibt es bei beiden Visualisierungen die Möglichkeit zur Interaktion mit einem Drop-Down-Menü, in dem das gewünschte zu betrachtende Genre ausgewählt werden kann und sich die visualisierten Daten entsprechend anpassen. Diese Filterung bleibt bei Umschalten des Plots bestehen und wird direkt auf den anderen angewandt, sodass eine Interaktion mit einem Plot zur Änderung des anderen führt. Durch die Anzeige des Menüs sowie die sich anpassende Unterüberschrift des Plots ist das gewählte Genre nachvollziehbar. Die Zielgruppe kann durch die Interaktion mittels genrefilter die Spiele und die Abhängigkeiten derer zwischen den Regionen sinnvoller vergleichen und vielversprechende Genres und in diesem möglicherweise Regionen für die Investition in weitere Marktstudien eingrenzen.

## 4 Implementierung

Im folgenden Kapitel wird die Implementierung der zuvor beschriebenen Visualisierungen und Interaktionen beschrieben. Der Code wird in fünf Elm-Module unterteilt, bestehend aus je einem Modul für die drei einzelnen Visualisierungen, einem für Datenstrukturen und einem zur Zusammensetzung und Ausführung des Hauptprogramms. *TreeHierarchy* dient als Ausgangspunkt der Anwendung, wobei *MainScatterParallel* den Hauptteil der Anwendung durch die eigentliche Verbindung der Visualisierungen enthält. Innerhalb der Visualisierungsmodule werden zuerst kleinere Funktionen definiert, die im späteren Verlauf genutzt werden, bis allgemeine Definitionen der Plots schließen. Über alle Module hinweg werden diverse Bibliotheken genutzt, die zu Beginn jedes Moduls importiert werden.

Insgesamt zeigte sich die Implementierung der Anwendung als zeitaufwendig, da trotz der umfangreichen Übungsaufgaben die Routine und teils das Verständnis im Programmieren mit Elm fehlte. Die Übungsaufgaben waren jedoch enorm hilfreich und dienten oft als Grundlage der Funktionen. Im gesamten Projekt gab es neue Erkenntnisse und Verständnis für die Funktionen der Sprache und des Quellcodes, sodass sich ein zusätzlicher großer Lerneffekt einstellte. Trotz der guten Vorlagen der Abgaben der Autorin zu den Übungsserien traten einige Schwierigkeiten auf, die in den folgenden Unterkapiteln Erwähnung finden.

### 4.1 Data.elm

Dieses Modul dient der Deklaration nahezu aller verwendeten Datentypen sowie der Implementierung der Decodierungen. Einzig *type Model* und *type Msg* sind aus Gründen der Übersichtlichkeit und Verständlichkeit des Codes in *MainScatterParallel* dort nicht vertreten.

Zunächst sind die Funktionen *decodeGameSales*, *csvString\_to\_data* und *gamesSalesList* zur Decodierung der CSV-Dateien implementiert. Die einzelnen Felder der CSV-Datei werden als String decodiert und in eine Liste vom Typ *GameSales* geschrieben. Darauf folgt die Decodierung der JSON-Datei mittels *treeDecoder*. Beide Decoder finden im *update* in *MainScatterParallel* zur initialen Überschreibung der Daten ihre Anwendung. Sie basieren den für die Übungen sieben und neun verwendeten Decodern und wurden an die hier verwendeten Daten angepasst.

Es folgt die Deklaration jener Datentypen, die für alle Visualisierungen außer die *TreeHierarchy* wichtig sind. *Type alias GameSales* definiert alle Daten der CSV-Datei wie sie durch den Decoder decodiert werden und später im Model in *MainScatterParallel* weiterverwendet werden. Die Deklarationen *type RegionType* und *type PlotType* als CustomType werden für die Auswahl der Achsen in den Visualisierungen 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. Es folgen die durch die Übungsserien bekannten Datentypen für die spezifischen Visualisierungen.

Zuletzt werden die für die Buttons wichtigen Funktionen zur Typkonversion von einem String zum jeweiligen CustomType und mit Ausnahme vom PlotType umgekehrt beschrieben. Dies ist nicht nur zur Erstellung von Buttons, die nur Strings erlauben, nötig, sondern auch zur bspw. korrekten Anzeige von aktuell ausgewählten Regionen in der Folge. 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

Für das Modul können große Teile des von der Autorin in Übungsserie neun erstellten Quellcodes verwendet werden und müssen lediglich minimal auf die verwendeten Daten angepasst werden. Begonnen wird 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 die Helferfunktion *convert* zur Konvertierung des Baumes basierend auf den Vorgaben zu Übung neun.

*Line* und *point* definieren die Linie zum Zeichnen der Verbindungen von Eltern zu Kindknoten bzw. zur Darstellung der Knoten an sich sowie die um 75° rotierte, fett gedruckte Beschriftung.

*TreePlot2* zur Zeichnung des Baumes ohne Scrollen und *treePlot* zur näheren Ansicht des Baumes werden folgend implementiert. Sie unterscheiden sich einzig in der prozentualen Vergrößerung der global definierten Höhen und Breiten um 200 bzw. 150 Prozent im *treePlot*. Auch die Plots an sich basierend stark auf dem Quellcode der Autorin zu Übung neun. Mittels lokaler Funktionen wird das Layout des Baumes bestimmt, bevor die Abhängigkeiten des Baumes mit Hilfe von *Dict.get* zum Erhalt der X- und Y-Werte berechnet werden, um die Pfade zwischen Eltern- und Kindknoten zu zeichnen. Die Funktionen *checkRootNegative* und *nodeValuePath* verhindern das fehlerhafte Zeichnen einer Linie zum nicht existenten Elternknoten der Wur-



zel. Nachfolgend werden die berechneten Verbindungen mittels Anwendung zuvor beschriebener Funktion *line* auf *nodeValuePath* gezeichnet. Berechnete Knoten werden durch Anwendung von *point* auf *nodeValue* gezeichnet. Die Definition des Designs inklusive Hovern wird im global definierten CSS *cssTree* vorgenommen.

In der *view*-Funktion werden lokal die Daten in einen Baum konvertiert, bevor dies neben dem CSS den beiden Zeichnungsfunktionen des Baumes übergeben und diese dann ausgegeben werden können. Die Anordnung sowie der Designstil der HTML-Elemente ist an *MainScatterParallel* angepasst.

Zuletzt werden global allgemeine konstante Einstellungen zum Plot definiert. Diese Definitionen basieren mit Unterschied bestimmter Werte auf den Vorgaben in der Vorlage zum Arbeiten mit Bäumen aus Übung neun und den Abgaben der Autorin zu Übung neun.

Die Implementierung der expliziten Baumhierarchie bereitete keine Schwierigkeiten, 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. Das CSS wurde gegensätzlich global definiert und das allgemeine Design Seite im *view* an jenes aus *MainScatterParallel* angepasst.

### 4.3 ParallelPlot.elm

Dieses Modul beginnt nach dem Import mit drei Funktionen zur Zuordnung der Daten zum benötigten Datentyp und der Reduzierung der Datensätze um solche, die Null-Values enthalten. Letzteres ist nach der manuellen Vorverarbeitung der Daten zwar nicht zwingend notwendig, jedoch kann so auf Fehler und Änderungen der Rohdaten eingegangen werden.

Die erste Hilfsfunktion dient mittels *Maybe.map2* dem Pipen. Basierend auf einer normalen *Maybe.map5* Funktion wird die zweite Hilfsfunktion um einen Parameter erweitert. Beide begründen sich aus den Dokumentationen der Elm-Core-Bibliothek zu *Maybe*. Da es kein *Maybe.map6* gibt, wird in *helpMapBig* 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 Achsentausch 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. Es werden unter anderem mittels lokaler Funktionen die Achsen in X-Richtung positioniert sowie folgend inklusive Beschriftungen in einem umgebenden Rechteck gezeichnet. Die Zeichnung der Datenpunkte wird durch *Shape.line* und *Shape.linearCurve* realisiert. Eine Änderung ist der beim Hovern über einen Datenpunkt anzuzeigende Text. Er wird lokal beschrieben, bevor dieser Funktion zuletzt die entsprechenden Daten übergeben werden.

Auch hier wird ein globales *cssParallel* definiert. Zur Entstehung des gewünschten Röntgen-effektes zur erleichterten Identifikation der Cluster und Auffälligkeiten der Datenpunkte wird die Opacity der Datenpunkte auf 0.5 festgelegt. Durch das Einfärben in einem stärkeren Grün-ton beim Hovern und dem Einblenden der Details zum Datenpunkt lassen sich die einzelnen Videospiele besser nachvollziehen.

Die Implementierung schließt mit den generellen Einstellungen für den Plot, welche mit Ausnahme bestimmter Werte auf den Abgaben der Autorin zu Übung sechs basieren.

Die Erstellung des *ParallelPlot* stützte sich 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 zuvor beschrieben. Hier war viel Ausprobieren und Verständnis, welcher Datentyp für welche weitere Funktion benötigt wird und welche durch die vorherige Definition der Datentypen vorhanden waren, nötig.

#### 4.4 Scatterplot.elm

Das Modul enthält ähnliche drei Funktionen mit demselben Zweck wie *ParallelPlot*. Die ersten beiden Funktionen gleichen sich. *FilterAndReduceGames* basiert jedoch auf den Abgaben der Autorin zu Übung eins. Die Daten werden zunächst mithilfe von *helpMapBig* vom Typ *GameSales* dem Typ *Maybe Point* zugeordnet. 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*, basierend auf den Abgaben der Autorin zu Übung vier, filtert mit

Eingabe einer Liste *GameSales* und des *RegionTypes* die Datenpunkte nach den je gewünschten Regionsauswahlen für die Achsen. Sie wird zur Interaktion innerhalb des Plots, zur freien Auswahl der Achsen und korrekten Anzeige der entsprechenden Datenpunkte benötigt. Die Anwendung wird in *MainScatterParallel* vorgenommen und beschrieben.

In *point* wird die Positionierung der Punkte sowie deren Beschriftung für die Weiterverwendung in *Scatterplot* festgelegt. Der *scatterplot* ist für das Zeichnen des Plots zuständig. Mit der lokalen Funktion *pointsXY* wird die Variabilität der Achsen und damit X- und Y-Werte der Punkte ermöglicht. 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 globalen Funktion *point* erreicht. Der *scatterplot* basiert auf den Abgaben der Autorin zu Übung 1.4 sowie Übung vier für die lokale Funktion *pointsXY*.

Global wird wiederum das *cssPoint* definiert. Zuletzt werden wie zuvor allgemeine, konstant für den Plot geltende Einstellungen, basierend auf Übung eins bis vier, global definiert.

Die Implementierung des Scatterplots verlief in weiten Teilen problemlos, wenn auch langsam durch ständiges Überprüfen der richtigen Datentypen. Die Umsetzung des Mappings und Filters nach Null-Werten stellte wie beim *ParallelPlot* ein Problem dar, wobei die Lösung vereinfachend größtenteils von dort übernommen werden konnte. Weiterhin musste beachtet werden, dass die Achsen des Plots frei wählbar, also variabel definiert sind, anders als 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 des Ursprungs von Funktionen, werden sie nicht dezidiert importiert, sondern nur das allgemeine Modul.

Es wird die gesamte Elm-Architektur verwendet, da das Modul 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 des *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 zuerst die initial anzuzeigenden Attributwerte mit Rücksicht auf die zuvor festgelegten Datentypen definiert und nach Anwendung der Decodierung in Success gespeichert. Weiterhin werden für alle 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.

Nachfolgend werden alle benötigten Button-Funktionen zur Überschreibung des Models bei Klick auf eine Variante implementiert. Wichtig ist die Nutzung der in *Data* definierten Umwandlungsfunktionen für alle außer den *buttonGenreType* von Strings, welche die Buttons benötigen, zu den entsprechend für das Model definierten Datentypen. Die Änderungen können dann in der Variante des *Msg*-Typen gespeichert und durch die *update*-Funktion das Model überschrieben werden. 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. Dem Anwender wird dadurch eine Filterung nach dem Attribut Genre ermöglicht.

In der *view*-Funktion werden die verschiedenen Fälle des Models zur Ansicht behandelt, wobei hier nur auf den Erfolgsfall näher eingegangen wird. Zunächst wird eine *let-in*-Konstruktion auf erster äußerer Ebene erstellt. In lokalen Funktionen wird *gameSalesData* sowie die Anzahl der Spiele insgesamt beschrieben. Zusätzlich wird hier der globale *filterGenre* zur Filterung der Daten nach gewünschtem Genre angewandt. Auch die Länge dieser Liste wird berechnet. Damit nun zwischen den Plots gewechselt werden kann, werden die verschiedenen Varianten bzw. Fälle des Datentyps *PlotType* wie im Record der Variante *Success* des Models aufgeführt im *in*-Teil angelegt. 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*-Ebenenkonstruktion sowie jeweils eine weitere *let-in*-Konstruktion in den Falldefinitionen. Diese befinden sich auf zweiter innerer Ebene, sodass die in der äußeren Konstruktion berechneten Zustände, sprich Genreauswahlen, auch hier gelten und bei einem Wechsel des Plots übernommen werden. In den lokalen Funktionen der inneren Konstruktion der Plots muss dazu die lokale *gameSalesDataFiltered* der äußeren Ebene angewandt werden. Würde in jeder der inneren Ebenen für jeden Plot die Filter nach Genre neu definiert werden und lokal angewandt statt in der übergeordneten Ebene, würden die Einstellungen respektive der Zustand des Models nicht übernommen werden.

Im ersten Case wird der ParallelPlot behandelt. Lokal wird die *assignmentAndReduce*-Funktion auf die gesamten Daten sowie auf die in der äußeren Ebene nach Genre gefilterten Daten angewandt. In *multiDimFunction* wird *multiDimenData* letztlich durch Übergabe der entsprechenden bereinigten und gefilterten Daten genutzt. Im *in*-Teil der inneren Ebene wird die Struktur und

das Design der Seite festgelegt. Für eine angenehmere und wenig ablenkende Ansicht werden die Farben den Grüntönen in den Plots angepasst. 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 Grüntonabstufungen 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. Eine transparente Anzeige der gewählten Einstellungen und verbleibenden Videospiele wird hinzugefügt. Schließlich wird dem *scatterplotParallel* das *cssParallel* sowie zwei Konstanten für die Höhe und die Aspect Ratio übergeben, bevor die zuvor lokal berechnete *multiDimFunction* folgt. Zuletzt wird noch ein Text mit Link direkt zurück zur TreeHierarchy in einem neuen Tab 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 den Informationstexten sowie der Auswahl der Achsen. Zuletzt wird dem *scatterplot* der *cssPoint*, die zuvor berechneten *gameSalesDataCleared* sowie die *valuesX* und *valuesY* übergeben. Zur Beschriftung der Achsen 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 unkompliziert umsetzbar. Besonders die Erstellung einer einheitlichen *update*-Funktion war jedoch zeitlich durch die Menge an Interaktionsmöglichkeiten aufwendig. Weiterhin schwieriger war die Implementierung der des Dropdown-Menüs für den ParallelPlot, da die Buttons zunächst je einzeln vorlagen. Durch Nachdenken und Ausprobieren vor allem mit der Funktion *multiDimenData* und verschiedenen Konvertierungsfunktionen konnten sie doch wie gewünscht realisiert werden. Zur Überprüfung immer wieder zeitaufwendig das *update* angepasst werden. Zunächst wurden beide Plots als untereinander auf einer Seite gezeichnet implementiert. Entsprechend war ohne Plot-Selektor nur eine *let-in*-Konstruktion nötig und die Filterung wurde für beide übernommen, was nicht dem Gewünschten entsprach. Nach schrittweisem Hinzufügen der Auswahl des Plots und allen zugehörigen Codeteilen, 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 potenzielle 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, bspw. 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 beinahe zu 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 befindet 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, ebenso 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. Auch 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 jedoch 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 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 Videospiele wäre auch durch bspw. Icontechniken möglich. Diese Methode wäre jedoch schwerer umzusetzen und aufwendiger zu analysieren. Der Grund liegt in der Darstellung aller Dimensionen eines Videospiele 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, abgesehen von Japan, bestätigt sich deutlich erkennbar die leicht bis vollständig positive Korrelation zwischen den jeweiligen Verkaufszahlen der Regionen. Im Scatterplot sind 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. 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.

Durch die Erkenntnisse aller 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 *Sports* zu konzentrieren. Dort sind sie selbst noch nicht stark präsent, es gibt weniger und in der Masse weniger starke Konkurrenz als bei *Action* sowie keine extremen Ausreißer. Weiterhin müssen nicht die Abhängigkeiten zwischen allen Regionen betrachtet werden, sodass eine Fokussierung und Abschöpfung eines Marktes Sinn ergibt. Die Fokussierung auf den europäischen oder nordamerikanischen Markt ist möglich und für einen 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 folgt eine knappe Literatursuche nach ähnlichen Anwendungen zu Videospielverkäufen oder Verkaufsdaten im Allgemeinen im Bereich der Informationsvisualisierung und Visual Analytics.

Der erste zu diskutierende Artikel ist *VizInteract: Rapid Data Exploration Through Multi-touch Interaction with Multi-dimensional Visualizations* von Chakraborty und Stuerzlinger.[2] 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 den Autoren bspw. 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 Datensatz der *Video Game*

*Sames 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, bspw. 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.[6] 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 Interessenbereich 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 Genres 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 Videospielemarkt 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.[6] 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.[2]

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

## Anhang: Git-Historie

### Literatur

- [1] Jacques Bertin und Wolfgang Scharfe. *Graphische Darstellungen und die graphische Weiterverarbeitung der Information*. Berlin: De Gruyter, 1982. ISBN: 978-3-11-006900-6. URL: <http://www.degruyter.com/doi/book/10.1515/9783110871494>.
- [2] 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.
- [3] Winnie Wing-Yi Chan. *A Survey on Multivariate Data Visualization*. 2006. URL: <https://people.stat.sc.edu/hansont/stat730/multivis-report-winnie.pdf> (besucht am 03.12.2022).
- [4] 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).

- [5] Michael Friendly und Daniel Denis. “The early origins and development of the scatterplot”. In: *Journal of the history of the behavioral sciences* 41.2 (2005), S. 103–130. ISSN: 0022-5061. DOI: 10.1002/jhbs.20078.
- [6] 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.
- [7] Alfred Inselberg und Bernard Dimsdale. “Parallel Coordinates for Visualizing Multi-Dimensional Geometry”. In: *Computer Graphics 1987*. Hrsg. von Tosiya L. Kunii. Tokyo: Springer Japan, 1987, S. 25–44. ISBN: 978-4-431-68059-8.
- [8] John Lamping, Ramana Rao und Peter Pirolli. “A focus+context technique based on hyperbolic geometry for visualizing large hierarchies”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95*. Hrsg. von Irvin R. Katz u. a. New York, New York, USA: ACM Press, 1995, S. 401–408. ISBN: 0201847051. DOI: 10.1145/223904.223956.
- [9] Dirk J. Lehmann u. a. “Visualisierung und Analyse multidimensionaler Datensätze”. In: *Informatik-Spektrum* 33.6 (2010), S. 589–600. ISSN: 0170-6012. DOI: 10.1007/s00287-010-0481-z.
- [10] o.A. *Videospiele: Weltweit*. 2022. URL: <https://de.statista.com/outlook/dmo/digitale-medien/videospiele/weltweit#umsatz> (besucht am 25.11.2022).
- [11] 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).
- [12] H-J Schulz, S. Hadlak und H. Schumann. “The Design Space of Implicit Hierarchy Visualization: A Survey”. In: *IEEE transactions on visualization and computer graphics* 17.4 (2011), S. 393–411. DOI: 10.1109/TVCG.2010.79.
- [13] Heidrun Schumann und Wolfgang Müller. *Visualisierung*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000. ISBN: 978-3-540-64944-1. DOI: 10.1007/978-3-642-57193-0.
- [14] 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).
- [15] 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).
- [16] John Q. Walker. “A node-positioning algorithm for general trees”. In: *Software: Practice and Experience* 20.7 (1990), S. 685–705. ISSN: 00380644. DOI: 10.1002/spe.4380200705.

- [17] Edward J. Wegman und Qiang Luo. “High Dimensional Clustering Using Parallel Coordinates and the Grand Tour”. In: *Classification and Knowledge Organization*. Hrsg. von Rüdiger Klar und Otto Opitz. Studies in Classification, Data Analysis, and Knowledge Organization. Berlin und Heidelberg: Springer, 1997, S. 93–101. ISBN: 978-3-642-59051-1.