

Projektbericht zum Modul Information Retrieval und
Visualisierung Sommersemester 2022

Marktanalyse des Videospielmarchtes

Analyse und Visualisierung der Verkaufszahlen der Videospiele auf der Plattform XBoxOne

Lena Arloth

25. November 2022

Inhaltsverzeichnis:

Inhaltsverzeichnis

1	Einleitung	2
1.1	Anwendungshintergrund	3
1.2	Zielgruppen	4
1.3	Überblick und Beiträge	5
2	Daten	5
2.1	Bereitstellung der Daten	7
2.2	Datenvorverarbeitung	7
3	Visualisierungen	9
3.1	Analyse der Anwendungsaufgaben	10
3.2	Anforderungen an die Visualisierungen	12
3.3	Präsentation der Visualisierungen	14
3.3.1	Visualisierung Eins	14
3.3.2	Visualisierung Zwei	15
3.3.3	Visualisierung Drei	16
3.4	Interaktion	17
4	Implementierung	18
4.1	Data.elm	18

4.2	TreeHierarchy.elm	19
4.3	ParallelPlot.elm	20
4.4	Scatterplot.elm	22
4.5	MainScatterParallel.elm	23
5	Anwendungsfälle	26
5.1	Anwendung Visualisierung Eins	26
5.2	Anwendung Visualisierung Zwei	28
5.3	Anwendung Visualisierung Drei	29
6	Verwandte Arbeiten	30
7	Zusammenfassung und Ausblick	33

1 Einleitung

Der weltweite Videospiegelmarkt über alle Plattformen hinweg ist mit einem Volumen von 156,17 Milliarde Euro im Jahre 2021 sehr groß und wird 2022 voraussichtlich auf 176,13 Milliarden Euro steigen.[5] 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 Daten der Vergangenheit. Eine Möglichkeit hierzu ist die Analyse der totalen Verkaufszahlen der Videospieletitel aufgeschlüsselt nach Genre weltweit.

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.

Als Problemstellung ergibt sich nun 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. Denn erst nach ihnen kann eine fundierte endgültige Entscheidung getroffen werden. Es ergibt sich also eine durchzuführende, grobe Marktanalyse der Verkaufszahlen der Videospiegelindustrie der Plattform *XBoxOne* 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 Potentialen.[3] Es ist für Unternehmen wie Publisher es sind 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 leiten sich folgende Fragestellungen 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?
 - 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, die mittels Informationsvisualisierung dargestellt werden können?
-

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 Potentiale und Risiken durch Korrelationen, Muster und Ausreißer beantworten können.[3] 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. 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.

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.[8] 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.[6] Somit behalten die Verkaufszahlen der *XBoxOne* ihre, wenn auch leicht verminderte, Aussagekraft.

Unzählige Publisher produzieren Spiele für die *XBoxOne* bzw. auch ihre Nachfolger und zählen viele Spiele unterschiedlicher Genres zu ihrem Repertoire. Dadurch ergibt sich der Sinn einer Übersicht. Zur besseren Vergleichbarkeit und Aussagekraft der Verkaufszahlen und ihrer Auffälligkeiten der Videospiele ist eine Filterung nach Genre sinnvoll.

1.2 Zielgruppen

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. Letzteres ist begrenzt auf Abteilungsleiter und fokussiert vor allem den oder die Leiter der Abteilung für Marketing sowie Forschung und Entwicklung. Das mittlere Management übernimmt dabei neben der Aufgabe des Mitentscheidens auch die des Präsentierens der Analysen und Ergebnisse, entsprechend der 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. Dazu zählt auch die Entscheidung, in welches Genre mit welchem Regionsfokus investiert bzw. weitere Studien durchgeführt werden sollen.

Daraus ergibt sich, dass es in der Zielgruppe kein detailliertes bzw. stark spezifiziertes Vorwissen zu speziellen Visualisierungstechniken gibt. Jedoch ist anzunehmen, dass sich vor allem das mittlere und obere Management mit Visualisierungen, die in der Betriebswirtschaftslehre häufiger zu finden sind, auskennt und diese ohne viele Erklärungen auswerten kann. Dazu zählen vor allem Balken- und Kreisdiagramme, Box Plots, Zeitreihendiagramme, Scatterplots, explizite Baumdiagramme und Parallele Koordinaten Plots. Bei spezielleren Visualisierungstechniken müssten mehr Erklärungen eingefügt werden sowie mehr Zeit zur Analyse und Entscheidung vorhanden sein. Dies wäre gerade für eine 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 harten Fakten, sprich konkreten Verkaufszahlen der einzelnen Spieletitel angesprochen. 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.

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.[7] Relevant zur Beantwortung der Fragestellungen sind dabei alle Attribute außer des Jahres der Veröffentlichung sowie der Tabellenposition. 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. Dies leistet einen Beitrag zur Beantwortung der übergeordneten Fragestellung und beantwortet die erste Unterfrage. Zur Beantwortung der zweiten Unterfrage nach 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 und damit auch zur Beantwortung der zweiten Unterfrage beitragen. 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 für das vorliegende Visualisierungsprojekt verwendete Datensatz stammt von der Plattform Kaggle von dem Nutzer *SID_TWR* mit dem Titel "*Video Games Sales Dataset*".[7] Die Daten entstanden laut Nutzer 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*.

Der Nutzer stellt drei Datensätze zur Verfügung, von denen für dieses Projekt der Datensatz *XBoxOne_GameSales* ausgewählt wurde. Hierbei ist zu beachten, dass die zuvor erwähnten Erweiterungen um Attribute stammend von *Metacritic* in diesem Datensatz nicht enthalten sind.

Der Originaldatensatz liegt als CSV-Datei vor und beinhaltet zehn Spalten mit 613 einzelnen Positionen, also Videospielen.

Zunächst ist eine eindeutige Identifikationsnummern gegeben, hier als *Positionen* bezeichnet. Weiterhin sind der Videospielnamen sowie das jeweilige Jahr der Veröffentlichung der Spiele aufgelistet. Das Attribut *Genre* kategorisiert die Videospiele in verschiedene Genre und *Publisher* ordnet jedem Videospiel seinen Verleger zu. Zuletzt sind die Attribute *North America*, *Europe*, *Japan*, *Rest of World* und *Global* aufgeführt. Sie stellen die Verkäufe der Videospiele in den Regionen Nordamerika, Europa, Japan, dem Rest der Welt und auf globaler Ebene dar. Die Einheit ist hier bei allen *millions of units*, also Millionen Stück verkaufter Kopien, was durch eine Recherche auf *VGChartz Video Games Sales* deutlich wird. Hochgeladen wurden die Daten circa vor drei Jahren. Der für das Projekt ausgewählte Datensatz bildet Videospiele von 2013 bis ungefähr 2020 ab, wobei seit dem Veröffentlichungsjahr 2019 keine Verkaufszahlen mehr eingetragen sind. Auch auf *VGChartz Video Games Sales* finden sich schon im Laufe des Veröffentlichungsjahres 2018 weniger bis keine Verkaufszahlen. Zwar wurden die Videospiele bis heute weiter eingetragen, jedoch ohne Aktualisierung der Verkaufszahlen. So ist davon auszugehen, dass die für das Projekt verwendeten Daten Verkaufszahlen ab 2013 bis circa 2018 abbilden. Der Stand der Zahlen hat sich auch laut *VGChartz Video Games Sales* aufgrund unbekannter Gründe nicht geändert. Es werden die kumulierten Verkaufszahlen in Millionen Stück in den Jahren seit Erscheinungsdatum bis circa Ende 2018 dargestellt.

Insgesamt werden die vorhandenen Daten als gut geeignet für die anvisierte Zielgruppe des oberen und mittleren Managements der Verleger von Videospielen für die Plattform Xbox One und das zuvor eingeleitete Zielproblem eingeschätzt. Die Daten ermöglichen eine grobe Übersicht über die verschiedenen Verleger und ihre Angebote in Form von Videospielen und bedienten Genres. Weiterhin schlüsseln sie detailliert auf, wie sich ein Videospiel seit seiner Veröffentlichung global, aber auch in den einzelnen Weltregionen verkaufte. Auf Basis dieser Daten ist eine Marktanalyse zur Lösung des Problems der fehlenden Übersicht und Kenntnis der Leistung der Videospiele eines Publishers am Markt sowie seiner direkten Konkurrenz im Genre möglich. Denkbar ist jedoch für eine weitergehende und tiefere Analyse die Einbeziehung von Daten zu Kritiken, Entwicklungskosten, Kosten pro Videospielkopie am Markt sowie die durch den Verkauf eines Spiels erzielten Umsätze. Diese Daten sind jedoch größtenteils nicht öffentlich zugänglich, unvollständig oder müssten mit sehr hohem Zeitaufwand aus diversen Quellen zusammengetragen werden. Die Visualisierung dieser Daten würde das hier gesetzte Zielproblem ergänzen, jedoch die anvisierte überblicksartige Marktanalyse überschreiten. In einer unabhängigen, weiteren Marktanalyse und den zugehörigen Visualisierungen wäre dies jedoch denkbar.

Der Datensatz enthält jedoch auch fehlende Werte bei den Verlegern, fortan Publisher genannt, und den Verkaufszahlen jeder Region. Die Daten werden also um jene Videospiele bereinigt, denen kein Publisher zugeordnet ist oder deren Publisher unbekannt ist. Weiterhin werden jene Videospieldaten gelöscht, deren Verkäufe in allen Regionen bzw. global einen Wert von Null aufweisen. Zuletzt wird die Zahl der Videospiele um jene reduziert, deren Publisher nur ein

Spiel herausgibt. Die Spalte des Erscheinungsjahres des jeweiligen Videospieles wird aus der hier durchgeführten Marktanalyse und ihrer Visualisierung ausgegliedert. 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. Eine Analyse dieser Daten gliedert sich in mögliche weiterführende Visualisierungen, wie im obigen Abschnitt beschrieben, ein.

2.1 Bereitstellung der Daten

Wie zuvor beschrieben, werden die originalen Daten über *Kaggle* zur Verfügung gestellt. Diese werden als CSV-Datei heruntergeladen und zusätzlich zur Bearbeitung 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. Die Dateien mit den Originaldaten enthalten die Endung *_Original*, Dateien mit Testdaten die Endung *_test* und finale für das Projekt nutzbare Dateien die Endung *_Projekt*. Die in Tabellenform überführten Originaldaten werden im ODS-Format bearbeitet, gelöscht, in einer neuen Tabellendatei gespeichert und wieder in CSV konvertiert. Die Datenvorverarbeitung im Detail ist im folgenden Unterkapitel einsehbar. Die CSV-Datei mit der Endung *_JSON* bildet die Grundlage für die Erstellung der JSON-Datei.

Um die Visualisierung mittels explizitem Baumdiagramm zu erstellen, wird eine JSON-Datei benötigt, in der die Beziehungen der Publisher, Genre und Videospiele zueinander beschrieben sind. Videospiele sind hier Kinder der Genres, die wiederum Kinder der Publisher sind und unter einem Wurzelknoten zusammengefasst werden. Für jeden Publisher sind die Genres als Kinder beschrieben, die er bedient. Unter jedem Genre, das ein Publisher bedient werden dann die Videospiele kategorisiert. Die Datei findet sich im Ordner *Daten/JSON* mit der Bezeichnung *XBoxOne_GamesSales_Projekt.json*.

Für die Realisierung der Visualisierungstechniken des Scatterplots und der parallelen Koordinaten wird dieselbe CSV-Datei benötigt. Sie ist im Ordner *Daten/CSV* mit der Bezeichnung *XBoxOne_GamesSales_Projekt* zu finden. In dieser Datei sind die nachfolgend detailliert beschriebenen Modifikationen enthalten. Sie beinhaltet entsprechend die alle Positionen, Videospielnamen, Genres, Publisher, Verkäufe in Nordamerika, Europa, Japan, dem Rest der Welt und die globalen Verkäufe. Dezimaltrennzeichen sind als Punkt realisiert und die einzelnen Werte in der Datei sind mit Kommata voneinander getrennt.

2.2 Datenvorverarbeitung

Zur besseren Nachvollziehbarkeit der Datenvorverarbeitung wird im GitHub Repository im Ordner *Daten* eine zweite README.md-Datei erstellt, die den Fortlauf der Vorverarbeitung dokumentiert. Da der originale Datensatz einige Anpassungen verlangt, um für die Visualisierung sinnvoll nutzbar zu sein, ist eine Datenvorverarbeitung und Modifikation unausweichlich. Die

benötigte Datengrundlage wird in sieben Schritten erreicht.

Zuerst werden die Daten als CSV-Datei heruntergeladen und unter der Bezeichnung *XBoxOne_GamesSales_Original.csv* im Ordner *Daten/CSV* gespeichert. Außerdem wird die CSV-Datei in ein ODS-Format konvertiert, um mit dem Programm *Open Office Calc* besser bearbeitet werden zu können. Diese Datei ist unter *XBoxOne_GamesSales_Original.ods* im Ordner *Daten/Tabelle* zu finden.

In einem zweiten Schritt werden Testversionen der Originaldaten mit den ersten 20 Positionen erstellt im CSV- und ODS-Format erstellt. Mit ihnen können die Visualisierungen zunächst übersichtlich auf Funktionalität getestet werden. Erst in einem späteren Schritt werden dann nur die Quellen für die Daten im Programm ausgetauscht. Die Testdaten werden zudem um die Spalte des Erscheinungsjahres reduziert. Eine Erklärung folgt in Schritt drei. Die Testdateien werden unter *XBoxOne_GamesSales_test.csv* und *XBoxOne_GamesSales_Original.ods* in den Ordnern *Daten/CSV* und *Daten/Tabelle* gespeichert. Im dritten Schritt erfolgt die eigentliche Modifikation in den Originaldaten in originaler Größe. Die Originaldatei im ODS-Format wird zusätzlich als *XBoxOne_GamesSales_Projekt.csv* im Ordner *Daten/Tabelle* gespeichert, sodass hier ohne Veränderung der Originaldaten modifiziert werden kann. Zunächst wird die Spalte *Year* gelöscht. Diese Daten sind für die Visualisierungen nicht nötig. Die Daten geben keine Grundlage für eine Visualisierung als Zeitreihe. Das Erscheinungsjahr könnte jedoch in Relation zu den Verkaufszahlen betrachtet werden, da diese im ersten Moment durch die Kumulation der Verkaufszahlen über die Jahre größer sein können, je länger das Spiel auf dem Markt ist. Jedoch kann auch ein bspw. erst 2017 erschienenes Videospiel höhere Verkaufszahlen aufweisen als eines, das bspw. im Jahr 2013. Zuverlässige, eindeutige und schnell erkennbare Schlüsse, die für eine hier angestrebte überblicksartige Marktanalyse nötig sind, können somit nicht ermöglicht werden. Es werden zusätzlich jene Positionen, also Videospiele, eliminiert, die einen Wert von Null in allen Regionen der Verkäufe enthalten. Auch unter Anbetracht der Tatsache, dass hier zumindest geringe Verkäufe unter 100.000 verkauften Stück erzielt worden sein könnten, ist ein Vergleich und eine Visualisierung sinnlos. Weiterhin werden die Positionen gelöscht, die keinen Publisher oder den Wert *Unknown* an dieser Stelle enthalten. Sie sind für die Zielgruppe und das Zielproblem irrelevant. Zusätzlich wird der Datensatz um die Positionen minimiert, die von einem Publisher stammen, der nur dieses eine Videospiel verlegt. Ein Vergleich der Spiele des Publishers untereinander und mit Konkurrenten in dem Genre ist nicht sinnvoll. Zudem ist damit zu rechnen, dass sich jene Publisher sehr spezifisch ausgerichtet haben und keine hier visualisierte überblicksartige Marktanalyse benötigen. Die entstehende Datei wird zuletzt wieder in das CSV-Format konvertiert und unter *XBoxOne_GamesSales_Projekt.csv* im Ordner *Daten/CSV* gespeichert.

Im vierten Schritt wird der Name des Publishers *Namco Bandai Games* in *Bandai Namco Games* vereinheitlicht, da beide denselben Verleger darstellen. Dieser wurde 2014 in letzteren umbenannt.

Als Vorbereitung für die Erstellung der JSON-Datei wird im fünften Schritt eine weitere

CSV-Datei erstellt, die die Abhängigkeiten für die gewünschte Baumstruktur als Spaltenüberschriften enthält. Die benötigten Daten des Publishers, Genres und Videospielnamens werden entsprechend der Struktur übertragen. Die Datei ist unter *XBoxOne_GamesSales_JSON.csv* im Ordner *Daten/CSV* zu finden.

Der fünfte Schritt beinhaltet die Konvertierung der erstellten CSV-Datei mit den Abhängigkeiten in eine JSON-Datei. Diese wurde mittels des Online-Tools *convertcsv.com* vorgenommen. Der entstandene Code wird im Editor *Visual Studio Code* in eine neue Datei kopiert und zunächst unter *Daten/JSON* gespeichert. Zusätzlich wird eine Testdatei im JSON-Format mit nur 20 Videospielen erstellt, um schneller und übersichtlicher entwickeln und das Entwickelte testen zu können. Auch diese Datei wird im entsprechenden Ordner gespeichert. Die weiteren Schritte werden sowohl für die Testdatei, als auch für die große, später genutzte JSON-Datei durchgeführt.

Zuletzt wird der JSON-Code auf Fehler in den Abhängigkeiten kontrolliert und verbessert. Dasselbe wird respektive in der Datei *XBoxOne_GamesSales_JSON.csv* vorgenommen. In der JSON-Datei wird ein Wurzelknoten hinzugefügt und die leeren Felder, die im expliziten Baumdiagramm leere Blattknoten darstellen würden, gelöscht. Zusätzlich wird pro Genre der jeweilige Eltern-Publisher wenn möglich abgekürzt in Klammern hinzugefügt. Somit können die Genres eindeutig identifiziert werden und es kommt nicht zu Überschneidungen der Pfade und mehreren eingehenden Pfaden pro Knoten respektive mehreren Elternknoten pro Kindknoten.

Für das in diesem Projekt anvisierte Ziel und den beschriebenen Anwendungsfall sind weitere Berechnungen nicht nötig. Desweiteren ist eine Umrechnung der Einheiten der Verkäufe nicht ratsam, da sie in der vorliegenden Form sehr übersichtlich sind. Bei einer Umrechnung in bspw. Tausend Stück würde die Zahl schon zu unübersichtlich und nicht auf den ersten Blick erkennbar sein, wie es für die Zielgruppe notwendig ist. Zur Erläuterung der Einheiten wird ein Informationstext erstellt und die Achsenbeschriftungen entsprechend angepasst.

Zunächst wurde eine Filterung der Daten auf Null-Werte in den Verkäufen jeder Region in Betracht gezogen, sodass diese Videospiele, sollten sie auch nur in einer Region einen Null-Wert enthalten, nicht angezeigt würden. Dies ergibt jedoch wenig Sinn im Zusammenhang einer übersichtsartigen Marktanalyse. Hierfür sind auch Informationen über keine erfolgten Verkäufe und die Region, in der nichts verkauft werden konnte, wichtig. Zu beachten ist jedoch, dass als Null-Werte auch solche aufgeführt sind, deren Verkaufszahlen bei unter 0.01 Millionen Stück verkaufter Videospieldkopen liegt. In Anbetracht der Größe der Märkte sind diese sehr geringen Verkaufszahlen jedoch als Null und entsprechend verbesserungswürdig anzusehen. Entsprechende Erklärtexte sind notwendig.

3 Visualisierungen

Im Folgenden Kapitel wird zunächst eine Analyse der Anwendung zur Lösung des Zielproblems einer überblicksartigen Marktanalyse der Verkaufszahlen von Videospielen der Plattform

XBoxOne durchgeführt anhand derer Anforderungen an die Visualisierungen abgeleitet werden. Schließlich werden die Visualisierungen und Interaktionen präsentiert.

3.1 Analyse der Anwendungsaufgaben

Das Zielproblem dieses Projekts ist eine überblicksartige Marktanalyse des Videospielemarktes der Plattform XBoxOne anhand der Verkaufszahlen der Videospiele global und in einzelnen Regionen der Welt. Ziel der Marktanalyse aus Sicht der Publisher ist eine Übersicht über das eigene Verlagshaus sowie die Konkurrenz, die eigenen Videospieleerfolge und Kenntnisse über den Markt zur Erfolgskontrolle vergangener Jahre und informierteren Ausrichtung des Publishers in der Zukunft. Hieraus ergeben sich drei Unterziele zur Lösung des Gesamtproblems, die nachfolgend analysiert werden.

Zunächst soll die Frage nach den angebotenen Videospielen der Publisher und ihrer Konkurrenz in einem ersten, eher qualitativen Überblick beantwortet werden. Hierbei liegt der Fokus aufgrund der gegebenen Daten und deren guten Übersichtsmöglichkeiten auf der Identifikation der von einem Publisher angebotenen Genres und der Videospiele, die jenen zugeordnet werden können. So lässt sich erkennen, welche Schwerpunkte der Publisher in der Vergangenheit bzgl. des Genres legte und wieviele Spiele er in welchem Genre anbot. Besonders für größere Publisher, die viele Videospiele und/oder viele Genres anbieten, ist dieser 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 eine Rolle spielen können. Aufgrund der Zielgruppe des oberen und mittleren Managements der einzelnen Publisher, kann es durchaus der Fall sein, dass u.a. besonders die Darstellung zur Lösung des ersten Unterziels auch Geschäftspartnern und anderen Stakeholdern vorgestellt werden muss, die weniger stark in die Einzelheiten und Strukturen des Verlags eingebunden sind. Weiterhin wird die Annahme getroffen, dass die Zielgruppe häufig und in unregelmäßigen Abständen den Überblick über ihre Videospiele, deren Genres und ihre Konkurrenz auf den ersten Blick benötigt, um sie diese Daten ins Gedächtnis zu rufen. Mental soll bei den Personen der Zielgruppe ein Modell der Hierarchie und Zugehörigkeit der Videospiele zu Genres und dieser Genres zu den Publishern entstehen, um eine Übersicht zu ermöglichen. Dieses mentale Modell soll sich auch in der Visualisierung zeigen. Durch eine Visualisierung der Eltern-Kind-Beziehungen, also der Hierarchie, ist eine Lösung der Fragestellung des ersten Unterziels möglich.

Das zweite Unterziel beschäftigt sich mit einem konkreteren Blick auf die einzelnen Videospiele innerhalb eines Genres in den verschiedenen Regionen der Welt im Vergleich. Hierbei soll die Frage beantwortet werden, wie sich die Videospiele eines Genres im direkten Vergleich in den Regionen der Welt verkauften. Gleichzeitig soll verglichen werden können, wie performant konkurrierende Videospiele des Genres waren. Für Publisher X soll erkennbar sein, welche sei-

ner Spiele in dem ausgewählten Genre in welchen Regionen erfolgreich waren und welche nicht. Durch den Vergleich mit der Konkurrenz kann er möglich Gründe für die Verkaufszahlen seiner Spiele feststellen. Wenn bspw. eines seiner Spiele in allen oder nur einzelnen Regionen nicht gut verkauft werden konnte, ein Spiel eines Konkurrenten aber in diesen Regionen aber gut, dann deutet das auf einen starken Konkurrenten hin. Sind alle oder die meisten Spiele der Konkurrenz in allen oder einzelnen Regionen besser verkauft worden, als das Videospiel von Publisher X, dann deutet dies auf einen gesättigten Markt des Genres mit starker Konkurrenz hin, in das sich möglicherweise weitere Investitionen nicht lohnen. Andersherum können sich auch Marktlücken pro Genre identifizieren lassen. Bei der Bearbeitung dieser Probleme helfen vor allem Übersichten und Vergleiche als mentale Modelle mit Visualisierungen von Daten auf Positionen auf Achsen und der Verbindung der Daten eines jeden Videospiels. 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. Sie sind mitunter nicht zwingend zur Lösung nötig, verbessern das Verständnis gerade in Bezug auf Schnelligkeit und Übersichtlichkeit aber ungemein.

Zuletzt stellt sich die Frage, wie gut sich die Videospiele im Vergleich zu anderen Spielen des Genres im genaueren Vergleich verkauften. Hierbei legt sich der Fokus auf die möglichen Korrelationen zwischen den einzelnen Regionen und den globalen Verkäufen, aber auch auf konkreter sichtbare Verhältnisse von Verkaufszahlen eines Videospiels eines Genres auch im Vergleich zur Konkurrenz. So lassen sich zum einen anhand konkreter sichtbarer Zahlen als in der zweiten Visualisierung der Erfolg eines Videospiel im Genre kontrollieren, aber zum anderen auch Kenntnisse zu neuen Positionierungen des Publishers für neue Videospiele ermöglichen. Gibt es bspw. viele Videospiele eines Genres, die in der Region Europa schlecht verkauft werden konnten, aber mittelmäßige bis gute globale Verkaufszahlen zeigen, so kann das auf eine Präferenz der Region für andere Genres deuten. Sollte es zudem einen oder weitere wenige Ausreißer geben, die besonders gute Verkaufszahlen in dieser Region und global haben, 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. Zur Unterstützung der einfachen Erkennung von Korrelationen ist ein Modell mit nur zwei Dimensionen sinnvoll. Die Visualisierung soll auch die im mentalen Modell gewünschten, möglicherweise entstehenden Cluster abbilden können. Für die Bearbeitung dieses Unterziels ist das entstehende räumliche mentale Modell der Abbildung der Daten auf Positionen sehr wichtig und notwendig, um vor allem die Korrelationen und Cluster zwischen den Regionen erkennen zu können.

Insgesamt gliedern sich die Unterziele von einer groben, überblicksartigen Analyse in und

zwischen Publishern in und zwischen Genres zu einer konkreteren, aber dennoch überblickartigen Analyse der Verkaufszahlen innerhalb der Genres zur Erfolgskontrolle und Vergleich mit der Konkurrenz bis hin zu einer detaillierten Analyse der Verkaufszahlen im Vergleich von nur zwei Regionen und auch auf Korrelationen und Clustern zwischen den Regionen.

3.2 Anforderungen an die Visualisierungen

Aus den zuvor analysierten Anwendungsaufgaben und Unterzielen leiten sich diverse Anforderungen an die Visualisierungen ab. Allgemein sollen alle Visualisierungen effizient, expressiv und angemessen sein. So ist bedingt durch die Zielgruppe 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 ergibt also wenig Sinn, besonders komplexe, tiefgründige Visualisierungen zu nutzen, die viele Ressourcen verbrauchen, wenn es gerade 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.

Zur Beantwortung des ersten Unterziels, welches sich mit den angebotenen Videospielen und Genres der Publisher und ihrer Konkurrenz beschäftigt, ist eine Darstellung eben dieser Daten notwendig. Es müssen alle Publisher, sowie die jeweils von ihnen abgedeckten Genres und angebotenen Videospiele überblicksartig dargestellt werden. Wichtig ist Darstellung, die von den jeweiligen Publishern ausgeht und dann spezifischer wird. Um die eigenen Angebote schnell erkennen zu können, sollte nicht nur die Anzahl und Art der Genres je Publisher dargestellt werden. Zur mentalen Gewichtung der abgedeckten Genres ist auch die Darstellung der Anzahl der ihnen zugeordneten Videospiele je Publisher nötig. Aufgrund des geforderten Überblicks und der konkreten Darstellung der Videospiele, müssen auch die Titel der Spiele abgebildet werden. Weiterhin ist es zur Ermittlung der möglichen Konkurrenz nötig, alle Publisher in einer Übersicht in derselben Art und Weise darzustellen. Neben den oben beschriebenen Anforderungen ist eine der wichtigsten jedoch die schnelle Übersicht, vor allem über das eigene Verlagshaus. Es muss ohne viel Nachdenken und viele Erklärungen sowohl für die Zielgruppe, als auch für Stakeholder des Verlages ersichtlich werden, welche Genres der Publisher anbietet und welche und wieviele Spiele unter diesen Genres kategorisiert sind. Auf dieselbe Art muss entsprechend auch die Konkurrenz leicht und schnell ablesbar sein. 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 des zweiten Unterziels 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, in welchen Regionen der Welt sich welche Spiele eines Genres gut verkauften und wo entsprechend nicht. Weiterhin muss in derselben Darstellung sichtbar werden, wie die konkurrierenden Spiele abschnitten. Somit ist es von Nöten, alle in den Daten gegebenen Regionen in einer Darstellung abzutragen, in der sie miteinander verglichen werden können. Für eine bessere Vergleichbarkeit der Daten soll nach Genre kategorisiert werden und dieses je nach Betrachtungswunsch ausgewählt werden. In der Visualisierung soll hauptsächlich ein Spiel in verschiedenen Regionen verglichen werden, jedoch auch Spiele untereinander. Entsprechend ist es für das Design wichtig, dass Werte eines Videospieles miteinander verbunden werden. Für ein einheitliches Verständnis soll dies auf dieselbe Weise mit den konkurrierenden Videospiele passieren. Es braucht deshalb aber eine deutliche Unterscheidung der Spiele untereinander, bspw. mittels Farbe, sowie eine Beschriftung mit den Titeln der Videospiele. 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, Korrelationen zwischen den Regionen der Videospieleverkäufe und den globalen Verkäufen zu erkennen. So kann erkannt werden, wie die Zusammenhänge zwischen bspw. sehr guten Verkaufszahlen einer Region zu den Verkaufszahlen global gegeben sind. Weiterhin sollen noch einmal präziser die Verkaufszahlen der Videospiele der Publisher in einem Genre im Kontext dargestellt werden. Hierzu sollen nur die globalen Verkaufszahlen und die einer auswählbaren weiteren Region aufgezeigt 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 Videospieles 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.

Allen spezifischen Anforderungen voran steht die Anforderung der leichten, möglichst intuitiven Erkennbarkeit der Zusammenhänge und der schnellen, übersichtsartigen Vergleichbarkeit, die für die eingangs beschriebene Zielgruppe notwendig ist.

Die eingesetzten Visualisierungstechniken sollten günstig sein und deshalb einfacher gestaltet. Gleichzeitig sollten sie durch die Notwendigkeit von schnellen Entscheidungen zügig erstellbar und ebenso schnell und einfach verständlich sein, damit auch die in den Visual Analytics unübten Personen der Zielgruppe ein möglichst breites und präzises Verständnis für die Marktsituation, Verkaufszahlen und Korrelationen zwischen den Regionen erhalten können. Wichtig war eine Darstellung eines hierarchischen Überblicks über die Marktteilnehmer, deren bedienten

Genres und Titel sowie das Erkennen von Mustern und Korrelationen in den Verkaufszahlen der Spiele eines Genres zwischen den Regionen der Welt, auch über mehrere Dimensionen hinweg. Zusätzlich sollten Ausreißer sowie detaillierte Informationen zu einzelnen Datenpunkten erkennbar sein.

3.3 Präsentation der Visualisierungen

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

3.3.1 Visualisierung Eins

Zur Erfüllung der Anforderungen und in der Analyse gefundenen Problemstellungen wird für die erste Visualisierung ein explizites Baumdiagramm gewählt.

In der Visualisierung gibt es einen Wurzelknoten aus dem hierarchisch die Kindknoten der Publisher abgehen. Für jeden Publisher gibt es wiederum Kindknoten, die aus den vom Publisher abgedeckten Genres bestehen. Zur besseren Übersicht und Vergleichbarkeit werden nur diejenigen Genres pro Publisher abgebildet, die dieser Publisher anbietet. Die Genreknoten werden für jeden Publisher in derselben alphabetischen 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 noch einmal Kindknoten, die die Titel der Videospiele, die dem entsprechenden Genre zugeordnet sind, tragen. So entsteht eine hierarchische Übersicht über das Angebot jedes Publishers, sodass dieser sowohl sich selbst, als auch die Konkurrenz in wenigen Blicken erfassen kann. 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.

Es gibt verschiedene Alternativen zu dieser Darstellung bestehend aus impliziten Baumdiagrammen, radialen Baumdarstellungen und hyperbolischen Bäumen. 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 der Zielgruppe und deren Stakeholdern liegt, wird diesem Aspekt eine hohe Gewichtung gegeben.

3.3.2 Visualisierung Zwei

Als zweite Visualisierungstechnik zur Lösung des in der Analyse gestellten Unterziels zwei und seinen Anforderungen 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. Die Wertebereiche der Attribute sind die Liniensegmente. Die Liniensegmente sind im hier vorliegenden Fall 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 eine solche Sequenz von Liniensegmenten, aber entsprechend immer anderen Attributwerten, 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.

Zur Erkennung des Videospieltitels und des Publishers, wird beim Hovern über ein Liniensegment beides eingeblendet. Zudem wird das entsprechende Segment mittels Farbe hervorgehoben, um das Segment eindeutig erkennbar zu machen. Durch diese Designentscheidungen werden die allgemeinen Anforderungen an Expressivität und Effektivität erfüllt.

Quantitative Daten werden zudem 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, die jeweils dieselbe Einheit besitzen erreicht. Auch die weiteren Anforderungen vor allem an die Vergleichbarkeit und allge-

meine Übersicht über die Verkaufszahlen der Videospiele in einer leicht verständlichen Art und Weise wird erfüllt.

Zusätzlich können leicht Cluster erkannt werden, wenn bestimmte Videospiele, bspw. eine Reihe in einem Genre, alle in allen Attributen, sprich Regionen, gute Werte aufzeigen. Auch dies bietet Aufschlüsse über den Markt und entsprechende Konkurrenz.

Die konkreten Zahlen sind anhand der Achsenbeschriftung zumindest ungefähr ablesbar, jedoch überwiegt wie gewünscht das visuelle Erkennen und Vergleichen. Aus der Anforderung nach einer Vergleichbarkeit mit zumindest groben Datenwerten leitet sich zudem ab, dass keine Techniken der Datentinte verwendet werden. Dies wird vor allem durch dieselben Einheiten jeder parallelen Achse, der Vergleichbarkeit dieser dadurch und damit der intuitiven Einordnung von *hoch* und *niedrig* entlang der Achsen möglich.

Auch zu den parallelen Koordinaten gibt es Alternativen. Eine dieser Alternativen sind 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 das dritte Unterziel und die abgeleiteten Anforderungen durch 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. Auch hier sind die Einheiten wieder gleich und somit gut vergleichbar. 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 Videospiele 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 eine genauere Betrachtung dieser erfolgen. Entsprechend kann auch die Konkurrenz noch einmal anders und möglicherweise genauer

je nach Regionseinstellung wahrgenommen werden. Durch die Einblendung der exakten Attributwerte ist auch diese Anforderung erfüllt.

Die eher klassische Darstellung als Punkte erleichtert das Erkennen von Punktwolken respektive 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. Dank der Filterung in Genre und der recht unkomplizierten Darstellung ist auch die Angemessenheit gegeben. Wie in allen drei Visualisierungen gilt hier auch die Angemessenheit durch Effektivität.

Alternativ bestehen in diesem Kontext vor allem drei weitere Visualisierungsmöglichkeiten. Diese sind Q- und QQ-Plots, Histogramme und Korrelationstabellen. QQ-Plots eignen sich zwar auch zur Darstellung von paarweisen Vergleichen und Entdeckung von Korrelationen und Ausreißern, jedoch sind sie weniger intuitiv, benötigen ein gewissen Vorwissen und sind nicht schnell erkennbar. Histogramme wiederum Eine weitere Alternative sind Korrelationstabellen.

3.4 Interaktion

Interaktion in Visualisierung 1 Tree: Highlight Knoten + Anzeigen Beschriftung Hovern -> Übersicht, Details ohne zu Überfordern; Klick auf Link zu detaillierten Plots Link zu detaillierten Plots (Grund: Übersicht zuerst und dann etwas getrennt den Rest) Interaktion in Visualisierung 2 Parallele Koordinaten: Auswahl der Regionen pro Achse -> kein Switchen, da durch so viele Möglichkeiten durch so viele Achsen unübersichtlich, aber trotzdem sowas sinnvoll weil deshalb freie Auswahl sodass Anwender schneller und einfacher das auswählen kann was er sehen will ohne alles einmal durchzuschalten highlight der Datenpunkte durch Farbe beim Hovern + Anzeigen Titel, Publisher, konkrete Werte beim Hovern Interaktion in Visualisierung 3 Scatterplot: Auswahl der x und y Achse mit Regionen, die verglichen werden sollen highlight Datenpunkte durch Farbe + titel, publisher, konkrete werte beim hovern (einheitlich mit visualisierung 2) Interaktion Verknüpfung 2 und 3: Auswahl der Plots mit Dropdown -> wenn Plot gewechselt wird und dann wieder zurück, bleiben die Achseneinstellungen erhalten (Bei tree Text + Link) in jedem Plot Möglichkeit unten direkt zurück zum Tree zu kommen da dies Ausgangsseite und Übersicht Auswahl Genre für jede Visualisierung möglich -> bei Klick auf anderen Plot wird diese Auswahl auch dort gleich übernommen bzw führt dort zu einer Änderung durch Filterung nach Genre -> weniger einstellen, schnelleres Vergleichen

4 Implementierung

Im folgenden Kapitel wird die Implementierung der zuvor beschriebenen Visualisierungen und Interaktionen beschrieben. Der Code wird durch Kommentare ergänzt, um eine Übersichtlichkeit und Erklärung zu einzelnen Funktionen zu ermöglichen. Der Code wird in fünf verschiedene Elm-Module unterteilt, um die vollen Möglichkeiten der Sprache zu nutzen, das Programm performanter zu machen und den Code für Außenstehende übersichtlicher zu gestalten. 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. Hier 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 Bibliotheken *elm/core* und *elm/browser* werden nur für MainScatterParallel importiert. *elm/http* wird von MainScatterParallel, Data und TreeHierarchy verwendet, *elm/html* von MainScatterParallel, TreeHierarchy und Scatterplot. Zur Decodierung von JSON-Dateien für TreeHierarchy wird die Bibliothek *elm/json* in Data genutzt, für die Decodierung von CSV-Dateien für Scatterplot und ParallelPlot in Data *lovasoa/elm-csv* und *ericgj/elm-csv-decode*. Die Bibliothek *elm-community/typed-svg* wird für das Zeichnen der Plots in Scatterplot, ParallelPlot und TreeHierarchy gebraucht ebenso wie die Bibliothek *gampleman/elm-visualization*. *avh4/elm-color* wird nur in ParallelPlot und TreeHierarchy genutzt. *elm-community/list-extra* und *folkertdev/one-true-path-experiment* sind nur für das Zeichnen des ParallelPlots wichtig. Zur Berechnung des Layouts von Bäumen wird *wasdacraic/elm-tree-layout* in TreeHierarchy genutzt und *zwilias/elm-rosetree* zum Speichern der Bäume in einer Datenstruktur in TreeHierarchy und Data importiert.

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 MainScatterParallel sind dort nicht vertreten. Dies ist begründet in der sinkenden Übersichtlichkeit und Verständlichkeit des Codes in MainScatterParallel, sollten eben genannte, recht umfangreiche Typen in Data deklariert sein und in MainScatterParallel referenziert werden müssen.

In Data sind zunächst alle für dieses Modul benötigten Importe zu erkennen, bevor die Decodierung der CSV-Dateien mit den Funktionen *decodeGameSales*, *csvString_to_data* und *gameSalesList* geschieht. Hier werden die einzelnen Felder der CSV-Datei als String decodiert und dann in eine Liste vom Typ *GameSales* geschrieben. Darauf folgt die Decodierung der JSON-Datei mittels der Funktion *treeDecoder*. Beide Decoder basieren auf denen 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 gespeichert, welche dann im Model in *MainScatterParallel* weiterverwendet werden. Die Deklarationen *type RegionType* und *type PlotType* als *CustomType* werden für die Auswahl der Achsen in 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 Datentypendeklarationen für den *Scatterplot*, den *ParallelPlot* und die *TreeHierarchy* definiert. Somit sind alle drei Visualisierungen modular zusammengesetzt.

Zuletzt werden für die Anwendung mit Buttons wichtige Funktionen zur Datentypkonversion von einem String zu *PlotType*, String zu *RegionType*, *RegionType* zu String und String zu 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 beispielsweise korrekten Anzeige von aktuell ausgewählten Regionen. Weiterhin ist noch eine Konversion von einem *RegionType* zu einem Tupel aus *GameSales* und Float für die Funktion *multiDimenData* in *ParallelPlot* notwendig.

4.2 TreeHierarchy.elm

Im Modul *TreeHierarchy* werden wie immer in Elm die Importe definiert. Aus Data werden nur die für die Baumdarstellung benötigten Funktionen importiert. Danach folgt die Funktion *main*, die die Bestandteile der Anwendung definiert.

Entsprechend dieser werden nun die Funktionen *init* und *update* beschrieben. Ersteres initialisiert die Anwendung und lädt die Daten unter Anwendung des Decoders für JSON-Dateien. Im *update* wird deklariert, dass im Falle eines korrekten Ladens von Daten der neue Baum angezeigt wird, im Falle eines inkorrekten Ladens eine Error-Nachricht. Es folgt eine Helferfunktion *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 beziehungsweise zur Darstellung der Knoten an sich benötigt und in *treePlot* und *treePlot2* angewandt. Sie definieren die Linie beziehungsweise 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 wird der *treePlot* verwendet. Der einzige Unterschied zwischen den Plots ist die prozentuale Ver-

größerung der global definierten Höhe und Breite um 200 bzw. 150 Prozent im SVG des *treePlot*. Die Plots an sich sind stark basierend auf dem Code der Autorin zu Übung neun. Mittels lokaler Funktionen wird zunächst das Layout des Baumes berechnet, bevor lokal die Abhängigkeiten des Baumes berechnet werden, um die Pfade zwischen Eltern- und Kindknoten zeichnen zu können. 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 zunächst lokal der konvertierte Baum sowie das Layout berechnet. In einem `Html.div` werden geschachtelt durch weitere `divs` Überschriften, Erklärungen, Links und Buttons zur Ausgabe implementiert. Die Anordnung sowie der Designstil ist an `MainScatterParallel` angepasst. An den entsprechenden Stellen in der Abfolge der gewünschten Texte werden die beiden Baumhierarchien gezeichnet, wobei *treePlot* einen Balken zum horizontalen Scrollen erhält.

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

Im Modul `ParallelPlot` werden wieder zuerst die Importe definiert, wobei aus `Data` nur die für die Parallelen Koordinaten benötigten Funktionen importiert werden.

Die folgenden drei Funktionen dienen dem 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. Sie basiert auf <https://package.elm-lang.org/packages/elm/core/latest/Maybe>. Die zweite Hilfsfunktion basiert auf einer normalen `Maybe.map5` Funktion, die um einen Parameter erweitert wird <https://package.elm-lang.org/packages/elm/core/latest/Maybe>. Da es kein `Maybe.map6` gibt, wird hier die vorherige Hilfsfunktion angewandt, um dieses Problem zu umgehen. So kann transformiert werden wie es eine `Maybe.map6` Funktion machen würde, wenn es sie gäbe. `AssignmentAndReduce` wiederum mapped schließlich lokal mithilfe der `helpMapBig` die Daten vom Typ `GameSales` zu `Maybe GameSales` und reduziert diese dann 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. Hier wurden die Achsen jedoch getauscht, was aufgrund flexiblerer und schnellerer Auswahl der Achsen in diesem Projekt geändert wurde. `MultiDimenData` benötigt als Eingabe eine Liste von `GameSales` sowie fünf `RegionTypes`. Diese sind 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. All dies begründet sich in der Definition der Recordfelder in *type Model* sowie dem für das *update* benötigten Datentypen (*RegionType*, *String*). Als Output muss die Funktion den Typ `MultiDimData` ausgeben, der benötigt wird, um die Daten dem `scatterplotParallel` zu übergeben und den Plot zeichnen zu können. Eine besondere Schwierigkeit bei der Konstruktion der Funktion lag im Umfang dieser und im Verständnis welcher Datentyp benötigt wird. 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. Weiterhin wird das umgebende Rechteck festgelegt, sowie die Achsen und ihre Beschreibung gezeichnet. Die Zeichnung der Datenpunkte wird mittels `Shape.line` und `Shape.linearCurve` realisiert. Zudem wird der beim Hovern über einen Datenpunkt anzuzeigende Text lokal beschrieben bevor die Funktion dann zuletzt mit den entsprechenden Daten angewandt wird.

Wie auch in der `TreeHierarchy` schon geschehen, wird auch hier 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 eine Art 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.map`⁶. 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

Das Modul Scatterplot beginnt mit der Definition der Importe. Aus Data werden nur die für den Scatterplot benötigten Funktionen importiert.

Folgende drei Funktionen verfolgen denselben Zweck wie im ParallelPlot. Die ersten beiden Funktionen sind dieselben wie im ParallelPlot. *FilterAndReduceGames* ähnelt *AssignmentAndReduce* stark, basiert aber auf den Abgaben der Autorin zu Übung 1. Hier werden die Daten zunächst auch mithilfe von *helpMapBig* gemapped, jedoch vom Typ `GameSales` zum Typ `Maybe Point`. Dieser wird später für die Funktion *point* und diese für den Plot an sich benötigt. Weiterhin wird wieder mithilfe vom `List.filterMap` nach Null-Values gefiltert. Die gemappten und 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. Ausgegeben wird eine Liste `Floats`. Diese Funktion ist nötig zur Interaktion innerhalb des Plots, um die Achsen frei wählen zu können und damit auch die richtigen Datenpunkte anzuzeigen. 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 festgelegt und als SVG Nachricht ausgegeben. So kann *point* im *scatterplot* aufgerufen werden.

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. In weiteren lokalen Funktionen wird die Abbildung bzw. Umrechnung der Daten auf SVG berechnet. 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*. Der

scatterplot basiert auf den Abgaben der Autorin zu Übung 1.4 sowie Übung vier.

Als letztes werden allgemeine Einstellungen, die konstant für den Plot gelten sollen, 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 Erkenntlichkeit aus welchen Modulen welche Funktionen stammen, werden keine Funktionen dezidiert importiert, sondern nur das allgemeine Modul.

Für dieses Modul wird wie in *TreeHierarchy* die gesamte Elm-Architektur verwendet, da es den Scatterplot und die Parallelen Koordinaten 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*. Diese könnten zwar auch in *Data* definiert werden, verweilen aber zugunsten der Übersichtlichkeit und schnelleren Verständnisses des Codes in diesem Modul. 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 Daten als Liste vom Typ *GameSales* gespeichert. Zur Ermöglichung der Interaktionen gibt es weiterhin Felder für das Genre als *String*, die Achsen eins bis fünf als *RegionType*, die Namen der Achsen eins bis fünf als *String*, die X- und Y-Achse als *RegionType* und den gewünschten Plot als *PlotType*. 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. Bei dem *buttonGenreType* ist dies nicht nötig, da die Variante von *Msg* auch einen *String* benötigt. 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. Mit dieser Funktion kann der Anwender nach dem Attribut *Genre* filtern und ein für ihn sehenswertes Genre auswählen. Im Prinzip sorgt diese Funktion, wenn später im *view* angewendet, dafür, dass ausgewähltes Genre und angezeigtes Genre übereinstimmen, also Gleichheit besteht.

In der *view*-Funktion werden nun 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 hier *gameSalesData* sowie die Anzahl der Spiele insgesamt beschrieben. Zusätzlich wird auf dieser oberen Ebene der globale *filterGenre* angewandt, mit dem die Daten so gefiltert werden, dass dem Anwender nur sein gewünschtes Genre angezeigt wird. 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 nun ausgewählt wurde und wie das Model durch das *update* überschrieben wurde, wird nun einer der drei Fälle angezeigt. Der Filter bzw. diese Interaktionsmöglichkeit nach Genre soll nun in allen Plots übernommen und angewandt werden, ohne dass bei jedem Plotwechsel das Genre neu ausgewählt werden muss. Der Zustand des Models mit dem gefilterten Genre soll also bestehen bleiben. 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 nun auf zweiter bzw. innerer Ebene, sodass die in der äußeren Konstruktion berechneten Zustände, also Genreauswahlen, auch hier gelten und auch 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 auch 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 repräsentative der Zustand des Models nicht übernommen werden.

Im ersten Case wird der *ParallelPlot* behandelt. Lokal werden die gesamten Daten nun mittels der *assignmentAndreduce*-Funktion gemapped und reduziert, sowie die entstehende Listenlänge berechnet. Auch die in der äußeren Ebene nach Genre gefilterten Daten werden auf dieselbe Art behandelt. In *multiDimFunction* wird nun *multiDimenData* letztlich angewandt 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 per Html und vielen verschachtelten divs und Attributeeinstellungen 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 Grüntonabstufungen genutzt. Wichtiges wird mit deutlichen Änderungen gekennzeichnet, alles andere zur impliziten Lenkung mit sogenannten *Just noticable differences* in der Farbe und im Rand. Neben den Informationstexten und Überschriften werden hier auch die Buttons als Dropdown 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* vom Typ *MultiDimData* 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 zum *ParallelPlot*. 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 hier 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* als String, die zuvor berechneten *gameSalesDataCleared* als *XYData* sowie die *valuesX* und *valuesY* als Liste von Floats übergeben. Um die Beschriftung der Achsen zu ermöglichen, werden diese noch 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 definiert werden. Auch dies geschieht wieder mit denselben Designs wie zuvor. Weiterhin besteht hier die Möglichkeit durch einen Link zum eigentlichen Plot des Baumdiagramms zu gelangen. Da für das Baumdiagramm der Genrefilter irrelevant ist, muss er weder hier noch in *TreeHierarchy* angewandt werden.

Die Implementierung dieses Moduls war einerseits leicht, andererseits kompliziert. Die zuvor jeweils in eigenen Anwendungen getrennten Visualisierungen mussten zusammengefügt werden. Dabei wurden alle Teile, die nur für den jeweiligen Plot genutzt werden, in den entsprechenden Modulen belassen. Alles konnte stückweise kopiert und in einheitlichen Funktionen zusammengesetzt werden. Besonders die Erstellung einer einheitlichen *update*-Funktion war zeitlich durch die Menge an Interaktionsmöglichkeiten aufwändig. Zudem war es anstrengend, alles gleichzeitig im Blick zu behalten. Weiterhin schwieriger war die Implementierung der Buttons für den *ParallelPlot*. Zunächst waren diese Buttons je einzeln und nicht als Dropdown-Menü wie gewünscht mög-

lich. Durch einige Denkarbeit, Ausprobieren vor allem auch 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. Die Implementierung der Interaktion über die Plots hinweg durch die Verschachtelung des *view* mit Definition und erster Anwendung des Genrefilters auf äußerer Ebene, der Einführung des Datentypen *PlotType* und der Anwendung im *update* und *view* war ein Zufallsfund. Zunächst wurden beide Plots so implementiert, dass sie untereinander gezeichnet wurden und es noch keinen Selektor für den Plot gab, entsprechend nur eine *let-in*-Konstruktion nötig war. Hier wurde die Filterung des Genres logischerweise für beide Plots übernommen, jedoch entsprach diese Darstellung nicht den Wünschen der Autorin. 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 und der Anwendung des Variantenwechsels für den Plot 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 von Videospielverkäufen dargestellt. Dabei werden die Visualisierungen aus der Sicht eines mittleren Managers des Publishers *505 Games* betrachtet, der diese und die Erkenntnisse aus ihnen dem oberen Management des Publishers sowie Stakeholdern des Unternehmens präsentiert. Mittels dieser Präsentation sowie der Erkenntnisse aus den Visualisierungen sollen Indikatoren für potentiell sinnvolle und lukrative Investitionen in neue Videospiele bzw. Fortsetzungen vorhandener Titel eines Genres gefunden werden. Zudem können Vorentscheidungen zur Aufstellung eines neuen Videospiels im Markt getroffen werden, wobei zwischen einer breiten, globalen Aufstellung oder der Konzentration auf bestimmte Märkte resp. Regionen entschieden werden kann. Kann ein solcher Indikator für Potenzial gefunden werden, so kann nach der Präsentation entschieden werden, ob weitere detailliertere Marktstudien in Auftrag gegeben werden sollen. Sollte das Potenzial dadurch bestätigt werden können, kann ein neues Videospiel im analysierten Genre in Auftrag gegeben werden.

5.1 Anwendung Visualisierung Eins

Öffnet der Manager in der Präsentation die Website mit dem Programm, zeigt sich zuerst die erste Visualisierung, also das explizite Baumdiagramm. Dank der großen Überschrift der gesamten Visualisierungsanwendungen, sowie der Unterüberschrift für das Baumdiagramm wissen er und die Präsentationsteilnehmer sofort, worum es geht und wo sie sich befinden. Der Informationstext hilft beim Verständnis des Diagramms. Mittels der farblichen 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 *JND*, also *Just Noticable Differences* gearbeitet. Sollte eine Person mit Farbsehschwäche oder genereller Farbenblindheit unter den Präsentationsteilnehmern sitzen, so ist dies durch die Abstufung der Grüntöne kein Problem. Ob die Farbe mit einer anderen verwechselt wird, spielt für das Verständnis keine Rolle und durch die Abstufung auch im Kontrast ist die Website und Visualisierung auch ohne Farberkennung möglich.

Im ersten Ausschnitt des Baumdiagramms kann der präsentierende Manager nun 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, *Platform*, *Role-Playing*, *Simulation* und *Sports* die wenigsten. 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.* viele Spiele anbieten und entsprechend 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 präsentierende mittlere Manager, dass nachfolgend der Fokus auf das Genre *Action*, sowie grob vergleichend *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. Somit sind sie für eine Voranalyse der Verkaufszahlen im Videospielmarkt wie hier angenommen, zu aufwändig.

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, wobei der Navigationsbereich mit Auswahl des gewünschten Plots und Auswahl des Genres mittels Drop-Down-Menü kräftiger gekennzeichnet ist. 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. Durch die Achsenskalierung lässt sich dies nicht nur durch weniger Einwohner in Japan begründen. 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 gut 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* stechen heraus. Weiterhin lässt sich erkennen, dass die Verkaufszahlen im Genre *Action* über alle Regionen hinweg stabil scheinen. Verkaufen sich die Spiele in Nordamerika, so verkaufen sie sich in Europa in ähnlicher Weise, genauso wie im Rest der Welt und global. Dies wird auch durch unterschiedliche Achsenbelegungen bestätigt. Die genaue Reihenfolge der Abhängigkeiten lässt sich jedoch nicht erkennen.

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. Der mittlere Manager kann so für seine Zuhörenden den Schluss ziehen, dass eine Konzentration in diesem Genre auf eine bestimmte Region wenig Sinn ergeben würde und eine breite Aufstellung nützlich ist. Dass viele Titel ähnliche Verkaufszahlen aufweisen und somit je von vielen verschiedenen Personen gekauft werden, deutet auf starke Konkurrenz im Genre hin. Ein neuer Titel sollte also sehr überzeugend sein oder eventuell doch in einem anderen Genre platziert werden. Durch die sich im unteren Drittel der Skala befindlichen Verkaufszahlen bestehen jedoch noch Möglichkeiten.

Zudem können in dieser Visualisierung Außreißer, die in allen Regionen deutlich höhere Verkaufszahlen aufweisen, als starke Konkurrenz identifiziert werden. Durch Hovern über einen Datenpunkt, also eine Linie, kräftigt und verdunkelt sich ihre Farbe zur besseren Erkennung des Verlaufes und über dem Plot werden die zum Datenpunkt gehörigen Informationen angezeigt. So können er und seine Zuhörer leicht nachvollziehen, dass der Ausreißer im Genre *Action* das Spiel *Grand Theft Auto* vom Publisher *Rockstar Games* ist. Es scheint also in allen Regionen

starke Konkurrenz zu sein. Sollte nach einer weiteren Bestätigung der Erkenntnisse im Scatterplot die Entscheidung zu weiteren Marktstudien in dem Genre getroffen werden, müsste anhand weiterer Einflussparameter 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. Die Vermutung liegt nahe, dass diese Spiele dort nicht angeboten wurden. Deutlich erkennbar ist jedoch, dass einige Spiele in Nordamerika deutlich 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. 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 aufwändiger 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 noch 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. Dort verkaufen sie sich vergleichsweise schlecht. Jedoch ist durch die Ansammlung der Datenpunkte in einer waagerechten, bzw. in umgekehrter Anordnung der Achsenbelegung senkrechten, Linie auch keine negative Korrelation ersichtlich. Die Titel im Cluster der positiv korrelierten Spiele scheinen sich jedoch gegenseitig positiv zu beeinflussen. Es kann durch etwas geringeren Verkaufszahlen in Nordamerika vermutet werden, dass die europäischen Verkaufszahlen die nordamerikanischen positiv 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 Alternative zu Scatterplots ist sinnvoll nicht gegeben, da sie die beste und am leichtesten zu verstehende Darstellung von Beziehungen zweier Attribute eines Datenpunktes zueinander sind. Aufgrund der Zweidimensionalität der Verteilungen würden auch Q- und QQ-Plots mit nur einer visualisierten Dimension gegen f-Werte bzw. gegeneinander zum Erkennen von Verschiebungen nicht die gewünschten Ergebnisse liefern.

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. Eine 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 zum nordamerikanischen Markt für seine Verkaufszahlen nutzen. 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. Prinzipiell ist die Anwendung auf Videospielverkäufe oder andere Verkäufe sehr ähnlich, da bei beiden z.B. nach Regionen, Artikelart bzw. Genre oder Hersteller bzw. Publisher gefiltert werden kann. Auch ähnliche Schlüsse können aus den Ergebnissen

gezogen werden.

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. Das von den Autoren entwickelte Programm dient der Multi-Touch Interaktion, um multidimensionale Datenvisualisierung schneller und einfacher konstruieren und mit ihr interagieren zu können. Dies soll für Unerfahrene sowie Experten im Thema nutzbar sein und als Tool die Datenexploration vereinfachen. Im Tool sind verschiedene Datenvisualisierungen implementiert, ähnlich wie im hier durchgeführten Projekt, die dann mittels Verschieben und Übereinanderschoben zu neuen Plots resultieren können. So können laut der Autoren z.B. zwei orthogonale Histogramme durch Übereinanderschoben einen Scatterplot kreieren. Forschungsziel des konkreten Artikels die Beobachtung des Nutzerverhaltens beim Anwenden von einfacheren touchbasierten Interaktionen auf einem Tablet und auf einem großen Touchdisplay. Es soll bewiesen werden, dass *VizInteract* einfach nutzbare und leicht erlernbare Interaktionen, Flexibilität und Performance im Interface sowie eine schnelle multitouch Komposition und Editierung von Datenvisualisierungen ermöglicht.

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 ist ähnlich zu 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 und welche durch die Anwendung im vorliegenden Bericht deutlich wird. 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 nicht aus einer Anwendung, der Analyse dieser und der Ableitung von Anforderungen heraus entwickelt, sondern seine Hauptmerkmale anhand derer geprüft. Weiterhin steht nicht die Aussage der Visualisierungstechniken im Fokus, sondern die Handhabung des Tools.

Der zweite Artikel *Implementation of Business Intelligence for Sales Data Management Using Interactive Dashboard Visualization in XYZ Stores* stammt von Akbar et al. und beschäftigt sich mit Thema.[1] In diesem Artikel wird der Wert durch valide Informationen und ein gutes Daten-

management von Unternehmen zur Unterstützung von Entscheidungen und Verbesserung der Effektivität und Effizienz in fokussiert. Dazu wird ein Prozess aus Problemidentifikation, Evaluation und Planung, Analyse, Data Warehouse Designprozess und ETL vorgeschlagen, wobei im Sinne dieser Arbeit nur der letzte Schritt der Auswahl und Anwendung der Visualisierung wichtig ist. Als Anwendungsfall wird der *XYZ Store* genutzt, der tägliche Gebrauchsgegenstände verkauft und dessen bisherige Applikation zur Datenverarbeitung nicht bei der Entscheidungsunterstützung hilft. Mittels Implementierung von Business Intelligence Applikationen soll im Artikel Abhilfe geschaffen werden, sodass beim Verständnis von Kundenverhalten, der Erstellung realistischer Ziele sowie der Erkennung von Gelegenheiten zum Crossselling unterstützt werden kann.

Die in vorherigen Schritten gesammelten und vorbereiteten Daten werden dann mittels des Tools *Public Tableau* in Plots visualisiert. Hier gibt es ein Dashboard mit elf verschiedenen Graphen, welche auf die gewünschten Funktionen zugeschnitten sind.

Aus dem Bereich der Visual Analytics nutzen die Autoren im Artikel Liniendiagramme in verschiedenen Formen, teils noch unterteilt in Artikelkategorien, Balkendiagramme und Blasenendiagramme. Das *Products Sales Dashboard* ist hier eine Sammlung der mit verschiedenen Visualisierungstechniken erstellten Plots, an dem alle Verkaufsdaten an einem Ort gesammelt aufgezeigt werden.

Gemeinsamkeiten zwischen dem Artikel und dem vorliegenden Bericht liegen in der Wichtigkeit der mittels Visual Analytics aufbereiteten Informationen über Verkaufsdaten für verschiedene Zielgruppen. Diese weisen hierbei starke Ähnlichkeiten auf, da das interne Personal, vorrangig die Entscheidungsträger der Unternehmen angesprochen und die Daten mittels Visualisierungen zu sinnvollen Informationen für diese weiterentwickelt werden sollen. Auch der Zweck der Visualisierungstechniken zur Entscheidungsunterstützung und Erkennung von Gelegenheiten ähnelt sich.

Unterschiede liegen zunächst im Umfang der Entscheidungsunterstützung. Während in dieser Arbeit nur die Visualisierung betrachtet wird, wird im Artikel ein ganzheitliches Konzept mit mehrere vorhergehenden Schritten vorgeschlagen. Die Visualisierung und Abwägung der Techniken spielt eine sehr untergeordnete Rolle. Weiterhin unterscheiden sich die Arbeiten in der Tiefe der Analyse, da der Artikel nicht auf genaue Analysen als Beweis der Zweckerfüllung seiner Visualisierungen eingeht, sondern eher das Gesamtprodukt vorstellt. Die dort in Grafiken gezeigten Visualisierungstechniken scheinen zudem keine Möglichkeiten zur Erkennung von Korrelationen und hierarchischen Beziehungen z.B. zwischen Produkten zu beinhalten wie es in vorliegender Arbeit der Fall ist.

Zuletzt sei noch *Intelligent Visual Analytics Queries* von Hao et al. genannt.[4] In diesem Artikel möchten die Autoren vor allem Analysten bei der Erkennung von Eigenschaften und Beziehungen in ausgewählten Mustern mittels ihrer *Intelligent Visual Analytics Query*, kurz *IVQuery* unterstützen. Der Fokus liegt dabei 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 Analysemethoden Charakteristiken und Beziehungen zu anderen Attributen und Datenpunkten identifiziert werden, wobei analytische und visuelle Methode kombiniert 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 schließlich 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 werden und die paarweise Korrelation der gewählten Attribute berechnet. 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 Daten abgebildet werden können statt nur zwei Dimensionen.

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, da sie in vorliegender Arbeit nur für zwei genutzt werden sowie für mehrere als eher weniger sinnvoll angesehen werden. 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 verschieben, um Muster und Korrelationen besser zu erkennen. Gleichzeitig fehlt ihm diese Flexibilität bei Hao et al. Wie schon in den anderen Artikeln wird weniger die Beantwortung einer Frage aus dem Anwendungsfall heraus oder die Analyse mittels der Visualisierungen behandelt, sondern das Tool allgemeiner 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. Zum einen wurde ein Überblick über den Markt, das eigene Verlagshaus und die Konkurrenz ermöglicht, der in Präsentationen und zur wiederkehrenden Ansicht und Eigenanalyse der Zielgruppen genutzt

werden kann. Zum anderen wurde mittels der Visualisierungen und sich zeigenden Muster und Korrelationen in den Verkaufszahlen zwischen den Regionen pro Genre eine informiertere Entscheidungsunterstützung über potenzielle künftige Investitionen in neue Titel oder Fortsetzungen in den entsprechenden Genres bzw. tiefere und teurere Marktstudien hierzu ermöglicht.

Die Visualisierungstechniken wurden vom Groben zum Detaillierten angeordnet. 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 Datenpunkte in einer Darstellung 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, interessanten 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. Zur nötigen Erkennung von Details zu Publishern, Titeln und konkreten Verkaufszahlen wurden interaktive Anzeigen dieser Informationen beim Hovern über die Datenpunkte in allen Visualisierungen ermöglicht.

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. Dies ist jedoch auch Teil einer angesprochenen weiterführenden, detaillierten Marktstudie. Interessant wäre eine weitere Unterteilung der Regionen sowie das Hinzufügen der bislang unter *Rest of World* zusammengefassten Regionen 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.[4] 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 gut umgesetzt werden.

Anhang: Git-Historie

Literatur

- [1] Ricky Akbar u. a. “Implementation of Business Intelligence for Sales Data Management Using Interactive Dashboard Visualization in XYZ Stores”. In: *2020 International Conference on Information Technology Systems and Innovation (ICITSI)*. IEEE, S. 242–249. ISBN: 978-1-7281-8196-7.
- [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] 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).
- [4] 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.
- [5] o.A. *Videospiele: Weltweit*. 2022. URL: <https://de.statista.com/outlook/dmo/digitale-medien/videospiele/weltweit#umsatz> (besucht am 25.11.2022).
- [6] 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).
- [7] 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 (besucht am 13.11.2022).
- [8] 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).