

Projektbericht zum Modul Information Retrival und Visualisierung

Sommersemester 2021

Visualisierung von verschiedenen Weindaten

Richard Brennecke

Matrikelnummer: 220233156

GitHub Repository: <https://github.com/RicBre/Elm-Projekt-WineInformation>

Inhaltsverzeichnis

Inhaltsverzeichnis	II
1. Einleitung	1
1.1 Anwendungshintergrund	1
1.2 Zielgruppen	3
1.3 Überblick und Beiträge	4
2. Daten	5
2.1 Technische Breitstellung der Daten	6
2.2 Datenvorverarbeitung	6
3. Visualisierung	7
3.1 Analyse der Anwendungsaufgaben	7
3.2 Anforderungen an die Visualisierungen	8
3.3 Präsentation der Visualisierung	9
3.3.1 Visualisierung Eins	9
3.3.2 Visualisierung Zwei	10
3.3.3 Visualisierung Drei	12
3.4 Interaktion	12
4. Implementierung	13
5. Anwendungsfälle	14
5.1 Anwendung Visualisierung Eins	14
5.2 Anwendung Visualisierung Zwei	14
5.3 Anwendung Visualisierung Drei	16
6. Verwandte Arbeiten	17
7. Zusammenfassung und Ausblick	18
8. Literatur	i
Anhang	ii
Git Historie	ii

1. Einleitung

Weltweit wurden im Jahr 2020 mehr als 260 Millionen Hektoliter (mhl) Wein produziert [11]. Umgerechnet sind dies 26.000.000.000 Liter (26 Mrd. Ltr.), von denen allein Deutschland 8,4 mhl hergestellt worden sind [11]. Dies ist jedoch nichts im Vergleich zu der führenden Weinproduktionsnation Italien, welche im selben Jahr mit 49,1 mhl rund das neunfache der deutschen Menge produzierte [11]. Allein aus diesen Zahlen sollte sich bereits erahnen lassen, dass Weine global beliebt sind und diese in den verschiedensten Ausführungen existieren. Aus dieser Vielfalt an Weinen wurde ein Datensatz zusammengestellt mit rund 21 Tausend Einträgen, welcher in dieser Arbeit genauer betrachtet wird. Ebenso sollen die Informationen innerhalb des Datensatzes so aufbereitet werden, dass eine Visualisierung dieser Daten möglich wird. Diese werden anschließend mithilfe von verschiedenen Visualisierungstechniken abgebildet.

Das Hauptziel ist dabei, diesen Datensatz so bereitzustellen, dass die unterschiedlichen Zielgruppen ihre individuellen Fragen rund um das Thema Weine beantworten können. Um dies zu erfüllen, sollten zwei andersartige Charakteristika des Datensatzes gegenübergestellt werden können. Somit wäre es beispielsweise denkbar, das PreisLeistungsverhältnis bei den unterschiedlichen Merkmalen herauszufinden. Weiterhin sollte es möglich sein, mehr als nur zwei Charakteristika Datensatzes zu vergleichen, um so ein besseres Verständnis für die Eigenschaften und Zusammenhänge des Weines zu schaffen. Außerdem sollten die Visualisierungen eine Möglichkeit der Zuordnung der Weine nach der Herkunft ermöglichen, umso beispielsweise überprüfen zu können, ob die verschiedenen Produktionsmengen mit den Weinarten eines Landes gleichmäßig mitwachsen. Diese Ziele werden anschließend in diesem Projektbericht mithilfe von drei Visualisierungstechniken bearbeitet.

1.1 Anwendungshintergrund

Die Visualisierungstechniken, die in diesem Projektbericht verwendet werden, sind die des Scatterplots, der Parallelen Koordinaten und der Baumhierarchie. Diese werden nachfolgend kurz vorgestellt, um ein besseres Verständnis für diese zu schaffen.

In einem Scatterplot werden die Werte für zwei verschiedene numerische Variablen durch Punkte dargestellt. Die Position der einzelnen Punkte auf der horizontalen und vertikalen Achse geben die Werte für einen Datenpunkt an. Scatterplots werden verwendet, um Beziehungen zwischen Variablen zu beobachten. So zeigen die Punkte in einem Scatterplot nicht nur die Werte der einzelnen Datenpunkte an, sondern auch Muster, wenn die Daten als Ganzes betrachtet werden. Dadurch wird es möglich, verschiedene Korrelationsbeziehungen zu identifizieren. Weiterhin ist es mithilfe eines Scatterplots managebar, verschiedene Daten in unterschiedliche Gruppen zu unterteilen, je nachdem, wie eng die Punkte zusammenliegen. So lassen sich Ausreißer oder Lücken in den Informationen erkennen. [16]

Bei den Parallelen Koordinaten handelt es sich um einen Ansatz, mehrdimensionale Daten zu analysieren. Dabei werden die Werte auf verschiedenen Geraden eingezeichnet, wobei jede Eigenschaft als eine Dimension dargestellt wird. Diese Daten werden über die verschiedenen Achsen miteinander verbunden. Die Stärke der parallelen Koordinaten liegt vor allem darin, mehrdimensionale Muster und Vergleiche zu tätigen. Um dies zu erreichen, ist es wichtig, nicht zu denken, dass Linien eine Codierung von Zeitreihen darstellt (und somit eine Veränderung des Wertes von Zeitpunkt a nach Zeitpunkt b erschließt). Stattdessen stellt eine Linie im Parallelen Koordinatensystem eine Verbindung einer Reihe von Werten dar. So lässt sich beispielshalber einfacher erkennen, ob sich Werte innerhalb des Durchschnittes befinden oder besondere Ausreißer sind. Weiterhin ist es denkbar, generelle Aussagen über die verschiedenen Werte zu treffen, ob beispielsweise einige Werte insgesamt größer sind als anderen die restlichen Werte. [9]

Die Baumhierarchie ist eine Darstellung von Daten, welche hierarchisch aufgebaut sind. Dabei beginnt diese bei einem Element und verzweigt sich dabei mindestens zweimal. Diese durch die Verzweigung entstanden Komponenten können sich wiederum sich aufspalten, um so immer weiter eine hierarchische Beziehung darzustellen. Das fertige Diagramm ähnelt dabei einen Baum mit seinem Stamm und den Ästen. [1]

Diese Visualisierungstechniken stellen die verschiedenen Daten des Datensatzes dar. Dabei sind diese größtenteils Eigenschaften von verschiedenen Weinen. Auf diese Besonderheiten wird nun kurz eingegangen, um so ein grundlegendes Verständnis für diese zu schaffen. Die Eigenschaften, welche in den Visualisierungstechniken verwendet werden, sind Alkohol, Trinktemperatur, Süße, Säuregehalt, Körper, Gerbstoff und Jahr.

Der Alkoholgehalt des Weines entsteht bei der Gehrung des Weins. Dabei handelt es sich um einen natürlichen Prozess, welcher abläuft, nachdem die Trauben zerquetscht worden sind. Je nachdem, wie süß und reif die Weintrauben während der Ernte gewesen sind, zieht dies einen höheren Zuckergehalt nach sich und hat steigenden Alkoholgehaltes des Weines zur Folge. Der Alkoholgehalt eines typischen Weines liegt dabei zwischen 9 und 14 Volumenprozent. Falls ein Wein eines höheren Alkoholgehaltes haben sollte, wurde bei der Produktion künstlich Zucker oder Alkohol hinzugefügt. Die Zugabe von Zucker geht dabei mit einer Minderung der Güteklasse des Weines einher, da der Qualitätsanspruch bei diesen niedriger ist. [2]

Auch die Trinktemperatur hat einen Einfluss darauf, wie Wein bei Verkostung schmeckt. Dabei besitzt jeder Wein seine eigene ideale Trinktemperatur (welcher meist auf dem Etikett des Weines vermerkt worden ist). Eine Faustregel ist dabei für junge Rotweine eine Trinktemperatur von 14-15 Grad, kräftige Rotweine hingegeben bei 15-17 Grad und schwere Rotweine bei 17-19 Grad. Falls Rotwein zu warm getrunken werden, kann der Geschmack nicht mehr richtig zur Geltung kommen und der Alkohol überwiegt in dem Wein. Darüber hinaus könnte es sein, dass der Wein im Hals brennt. Bei Weißweinen hingegen liegt die Temperatur niedriger als bei den Rotweinen. So besitzt ein junger Weißwein eine Trinktemperatur zwischen neun und elf Grad, ein reifer Weißwein sollte bei elf bis 13 Grad getrunken werden. Falls ein Weißwein zu kalt getrunken werden sollte, überwiegt dabei die Säure und die restlichen Aromen des Weines gehen verloren. Da jeder Geschmack individuell ist, ist es auch möglich, mit der verschiedenen Trinktemperaturen des Weines zu experimentieren. Als Faustregel gilt dabei, sobald ein Wein kühler getrunken wird, verstärken sich die Säuren und Gerbstoffe, wobei sich die Aromatik abschwächt. Wenn jedoch die Trinktemperatur erhöht wird, kommt die Aromatik, Körper, Süße und der Alkohol des Weines mehr zum Vorschein. Wenn der Wein serviert wird, ist darüber hinaus besser, ihn tendenziell etwas kühler als die optimale Trinktemperatur zu servieren, da dieser sich schnell nach dem Ausschanken um zwei bis drei Grad erwärmt. [12]

Bei der Süße des Weines handelt es sich um eine der wichtigsten Bestandteile des Weines und verleiht ihm mit anderen Komponenten erst den Geschmack. Die Süße beschreibt dabei den Restzucker innerhalb des Weines. Dieser kann jeweils nach Anbaugebiet und Rebsorte auch unterschiedlich ausfallen. Dabei gilt jedoch primär, je reifer die Trauben sind, desto mehr süßer ist der Wein. Innerhalb des EU-Weingesetzes ist festgelegt, dass ein süßer (oder auch lieblicher) Wein mindestens einen Restzuckergehalt von 45 Gramm pro Liter aufweisen muss. Darüber hinaus ist es möglich, diesen im Nachhinein zu verstärken, indem Zucker hinzugefügt wird, um die natürliche Süße zu erhöhen. Dieser Vorgang ist als „Aufspritzen“ bekannt. [5]

Die Säure innerhalb eines Weines soll ihm die Frische und Eleganz verleihen. Diese entsteht beim Reifungsprozess an der Rebe. Dabei steigt der Zuckergehalt der Traube an, während der Säuregehalt sinkt. Das Verhältnis von dieser zu und Abnahme ist dabei nicht immer gleich. So können kühle

Nächte den Säurerückgang verzögern. Im Wein sind besonders zwei verschiedene Säurearten anzutreffen, welche die Weinsäure und die Apfelsäure sind. Dabei ist diese eine erwünschte Säure, da diese den Wein weich und angenehm schmeckend macht. Die Apfelsäure hingegen fördert, dass der Wein kantig und hart wird. Dies kann bei Weißweinen zum Teil erwünscht sein, bei Rotweinen jedoch wird diese in Milchsäure umgewandelt. Darüber hinaus ist der Apfelsäuregehalt auch ein guter Indikator für die Güte des Jahrganges, da dieser von der Witterung abhängt. So wird weniger Apfelsäure in sonnigen Jahren produziert und in kühlen Jahrgängen mehr. [14]

Bei dem Körper des Weines handelt es sich um eine Möglichkeit, den Eindruck des Weines auszudrücken, welcher dieser im Mund hinterlässt. Der Körper beschreibt somit die Gerbstoffe, Restsüße und die Säure des Weins. Dabei wird in einen leichten, mittleren und voluminösen Körper unterschieden, was zu der Bezeichnung leichtgewichtigem, mittelgewichtigen und schwergewichtigen Wein führt. So sind vor allem die Gerbstoffe für den Körper verantwortlich, da meistens ein höherer Gerbstoffgehalt zu einem voluminöseren Körper führt. Aber auch wenn der Alkohol kein Bestandteil des Körpers ist, spielt dieser beim Körper eine Rolle, denn je spürbarer der Alkoholgehalt des Weines ist, desto voluminöser erscheint dieser dann im Mund. [4]

Die Gerbstoffe (oder Tannine genannt) kommen primär in Wein vor, bei dem die ganze Weintraube verarbeitet wird. Dies liegt daran, dass die meisten Gerbstoffe in dem Kern der Traube und in der Schale vorzufinden sind. Dementsprechend besitzt ein Rotwein mehr Gerbstoffe als ein Weißwein, da hierbei die gesamte Weintraube verarbeitet wird. Jedoch können auch bei der Gärung in Holzfässern weitere Gerbstoffe hinzukommen. Beim Trinken des Weines rufen die Gerbstoffe den herben bis bitteren Geschmack und das Gefühl, dass sich die Mundschleimhäute zusammenziehen oder der Mund trocken ist, hervor. Je nach Stärke des Weines sind die Gerbstoffe im hinteren Bereich des Mundes zu spüren, bei besonders kräftigen Gerbstoffen auch im gesamten Mund. Zu viele Gerbstoffe geben den Wein ein eher starkes, herbes oder bitteres Aroma, zu wenig Gerbstoffe führen jedoch zu einem flachen Geschmack. Bei der Lagerung kann es darüber hinaus sein, dass sich die Gerbstoffe Moleküle zu größeren zusammenfügen und somit dem Wein seinem besonderen Eigengeschmack verleihen. [3]

Ein großer Unterschied von Wein zu vielen anderen Getränken ist, dass dieser weiterhin in der Flasche reifen kann. Dies bedeutet, dass Weine in der Lage sind, mit der Zeit an Qualität zu gewinnen. Jedoch haben die meisten Weine ihren Trinkhöhepunkt schon nach einigen Jahren ihrer Abfüllung erreicht und sollten deswegen nicht übermäßig lange gelagert werden. Da ansonsten mit der Zeit Sauerstoff an dem Verschluss deines Weines vorbeikommt und im schlimmsten Falle Essig bildet. Dabei hängt es vom Gerbstoffgehalt des Weines ab, wie viel Sauerstoff er verträgt. Dieser macht die Gerbstoffe weicher und aufgrund dessen kann sich eine gute Frucht innerhalb des Weines entwickeln. Somit eignen sich Rotweine mehr zum Lagern als Weißweine, da diese weniger Gerbstoffe beinhalten. Beispielsweise ist einer der ältesten Weine ein 1870er Château Lafite ein Rotwein. [13]

1.2 Zielgruppen

Für die unterschiedlichen Visualisierungstechniken gibt es drei verschiedene Zielgruppen, welche potenziell an diesen interessiert sein könnten. Diese Gruppen sind: Weininteressierte, Weineinkäufer (-verkäufer) und der Weinexperte. Auf diese Zielgruppen wird nachfolgend genauer eingegangen.

Die Weininteressierten sind eine Gruppe, welche gerne Wein trinkt, jedoch kein bis kaum Vorwissen zu dieser Materie besitzt. Deswegen könnten diese mithilfe von verschiedenen Darstellungsweisen neue Erkenntnisse rund um das Thema der Weine gewinnen. Somit wäre es möglich, dass diese

Zielgruppe mithilfe dieser Visualisierungen bisher unbekannte Weine für sich entdeckt oder getrunzene Weine besser einordnen kann.

Bei der Gruppe der Weineinkäufer oder Weinverkäufer handelt es sich um die Zielgruppe, welche das meiste Wissen über Weine besitzen sollte. So muss diese ihre Kunden beraten und hervorragende Weine zu entsprechend zu angenehmen Preisen finden. Somit ist ihr Vorwissen sehr gut bis exzellent. Diese Visualisierungen sollten es dieser Zielgruppe ermöglichen, schneller neue Sorten zu entdecken, welche in das Sortiment passen. Weiterhin könnten diese Darstellungsweisen zur Kundenberatung verwendet werden, um den Kunden noch individueller und prompter zu beraten. So könnten beispielsweise neue oder außergewöhnliche Sorten für einen speziellen Klienten entdeckt werden.

Die Weinexperten sind eine Gruppe von Leuten, welche gerne Wein trinken und sich bereits mit diesem Thema auseinandergesetzt haben. Dementsprechend sollte ihr Wissenstand durchschnittlich bis gut sein. Diese Gruppe könnte mithilfe der Visualisierungen neue Sorten für sich entdecken, welche ihrem Geschmack entspricht.

1.3 Überblick und Beiträge

In diesem Projekt wurden die Daten von der Webseite „Kaggle“ verwendet. Dabei handelt es sich um Informationen rund um das Thema Wein. Diese Daten enthalten dabei Auskünfte über den Namen, Produzent, Herkunft, Typ und Verwendung des Weines. Darüber hinaus gibt es Angaben zu den verschiedenen Merkmalen des Weines wie Alkoholgehalt, Trinktemperatur, Süße, Säure, Körper, Gerbstoffe, Preis, Jahr und Größe einer Flasche. Genauere Informationen, was diese Eigenschaften ausdrücken, ist im Kapitel Das Hauptziel ist dabei, diesen Datensatz so bereitzustellen, dass die unterschiedlichen Zielgruppen ihre individuellen Fragen rund um das Thema Weine beantworten können. Um dies zu erfüllen, sollten zwei andersartige Charakteristika des Datensatzes gegenübergestellt werden können. Somit wäre es beispielsweise denkbar, das PreisLeistungsverhältnis bei den unterschiedlichen Merkmalen herauszufinden. Weiterhin sollte es möglich sein, mehr als nur zwei Charakteristika des Datensatzes zu vergleichen, um so ein besseres Verständnis für die Eigenschaften und Zusammenhänge des Weines zu schaffen. Außerdem sollten die Visualisierungen eine Möglichkeit der Zuordnung der Weine nach der Herkunft ermöglichen, umso beispielsweise überprüfen zu können, ob die verschiedenen Produktionsmengen mit den Weinarten eines Landes gleichmäßig mitwachsen. Diese Ziele werden anschließend in diesem Projektbericht mithilfe von drei Visualisierungstechniken bearbeitet.

Anwendungshintergrund zu finden. Diese Daten werden anschließend anhand ihrer verschiedenen Merkmale in einem Scatterplot, Parallelen Koordinaten und einer Baumhierarchie dargestellt. Wie diese Diagramme aufgebaut sind, wird im Kapitel Das Hauptziel ist dabei, diesen Datensatz so bereitzustellen, dass die unterschiedlichen Zielgruppen ihre individuellen Fragen rund um das Thema Weine beantworten können. Um dies zu erfüllen, sollten zwei andersartige Charakteristika des Datensatzes gegenübergestellt werden können. Somit wäre es beispielsweise denkbar, das PreisLeistungsverhältnis bei den unterschiedlichen Merkmalen herauszufinden. Weiterhin sollte es möglich sein, mehr als nur zwei Charakteristika des Datensatzes zu vergleichen, um so ein besseres Verständnis für die Eigenschaften und Zusammenhänge des Weines zu schaffen. Außerdem sollten die Visualisierungen eine Möglichkeit der Zuordnung der Weine nach der Herkunft ermöglichen, umso beispielsweise überprüfen zu können, ob die verschiedenen Produktionsmengen mit den Weinarten eines Landes gleichmäßig mitwachsen. Diese Ziele werden anschließend in diesem Projektbericht mithilfe von drei Visualisierungstechniken bearbeitet.

Anwendungshintergrund genauer erklärt. Dabei ist es möglich, mithilfe des Scatterplots zwei verschiedene Merkmale des Weines zu vergleichen, um entsprechende Muster zwischen diesen

beiden Eigenschaften zu erkennen und identifizieren. Dies könnte zu neuen Erkenntnissen rund um diese Kriterien führen. Mithilfe der Parallelen Koordinaten können mehr als zwei Eigenschaften miteinander verglichen werden. Dabei können verschiedenartige Trends und Besonderheiten bei den verschiedenen Merkmalen des Weines erkannt werden. Bei der Baumhierarchie wird es möglich sein, unterschiedliche Weine ihren Regionen zuzuordnen. Diese Informationen kann dann wiederum mit anderen Daten abgeglichen werden, um einen weiteren Erkenntnisgewinn zu ermöglichen.

2. Daten

Die Originaldaten, welche auf der Plattform "Kaggle" zu finden sind, wurden von einem Nutzer von einer koreanischen Webseite (welche nicht genauer angegeben worden ist) gesammelt und bereitgestellt. Diese Originaldatei beinhaltet 32 Spalten mit insgesamt 21.605 Datensätze. Da diese Daten teilweise koreanische Symbole enthielten wurde eine zweite Datei angelegt, um die entsprechenden Zeichen herauszufiltern und den Datensatz somit zu bereinigen. Dieses Dokument heißt „cleasingWine.csv“ und besitzt 31 Spalten und 21.600 Datensätze. Auf dieser Grundlage wurden die nachfolgend beschriebenen Datenbearbeitungsschritte vorgenommen. [7]

Der Datensatz von CleasingWine beginnt mit der „wineID“, welche eine einfache Indexnummer darstellt. Nach dieser Spalte folgt die Spalte mit dem Namen „name“, in dieser sind die Benennungen der einzelnen Weine eingetragen. Die Produzenten dieser Weine sind in der darauffolgenden Spalte der „producer“ zu finden. Aus welchem Land dieser Wein stammt, wird anschließend in der Spalte der „nation“ beantwortet. Anschließend folgen fünf Spalten mit dem Namen „local“ und der entsprechenden Nummer, welche die Herkunftsregionen angeben. Danach finden sich die Spalten „varieties“ und die dazugehörige Zahl (bis zwölf), um die Weine ihrer passenden Sorte zuordnen zu können. Nach diesen Spalten folgt die Spalte „typ“, welche den entsprechenden Typ des Weines enthält. Anschließend wird in der Spalte „use“ die Verwendung des Weines genauer festgelegt. Auf diese Spalte folgt die der „abv“, welche den Alkoholgehalt pro Volumen darstellt. Danach wird in der Spalte der „degree“ die optimale Trinktemperatur des Weines festgehalten. Nachfolgend wird mit den Spalten „sweet“, „acidity“, „body“, „tannin“, „price“, „year“, „ml“ die jeweilige Süße, Säure, Körper, Gerbstoff, Preis, Herstellungsjahr und Größe der Flasche des Weines definiert. Dabei sind ab der Spalte der „nation“ Nullwerte in der Datei enthalten. Zahlenwerte sind aber der Spalte „abv“, welche in diesem Datensatz zu finden. [7]

Die Datenqualität dieser Daten eignet sich für die Zielgruppe der Weininteressierten gut. Da diese eine angemessene Tiefe bieten, um neue Weine kennenzulernen und auch Zusammenhänge zwischen den Weinen Eigenschaften zu erkennen. Für die Gruppen der Weineinkäufer oder Weinexperten ist die Datenqualität ausreichend. Da hierbei vor allem durch die vielen Nullwerte keine vollständige Datenbasis vorhanden ist. Somit ist nicht garantiert das für jeden Wein, der ggf. entdeckt wird, die entsprechenden Daten parat sind, um eine fundierte Entscheidung treffen zu können. Weiterhin ist die Tiefe der Werte für diese Gruppen nicht ausreichend. Dies kommt daher das die Eigenschaften der Weine (Süße, Säure, Körper und Gerbstoffe) nur auf einer Skala von ein bis fünf angegeben werden. Dies ist eine Ungenauigkeit, welche von der Zielgruppe unerwünscht ist.

Jedoch sollte die Datenqualität ausreichend sein, um neue Erkenntnisse zu erhalten.

Dementsprechend reichen dieser Datensatz aus, um die verschiedenen Fragestellungen dieser Projektarbeit zu beantworten. Dabei sollte jedoch nicht zu viel Wert auf die Genauigkeit der Daten gelegt werden, da diese teilweise ungenau sind.

Darüber hinaus wurde für das Baumdiagramm eine weitere Datei erstellt, in welcher die Länder der Weine um Kontinente und deren Einteilungen ergänzt worden sind. Dies dient der bessern hierarchischen Darstellung der Weine und ihrer Nationen.

2.1 Technische Bereitstellung der Daten

Die technische Bereitstellung der Informationen erfolgt mithilfe des GitHub's Repository, in welchem das Visualisierungsprojekt umgesetzt worden ist. Dabei sind die Werte innerhalb des Ordners „Daten“ zu finden. In diesem sind unter dem Unterordner „Quelldaten“ befinden sich die ursprünglichen Angaben, welche von der Plattform „Kaggle“ heruntergeladen werden konnten. Eine genauere Beschreibung dieser Informationen ist im Kapitel Daten nachzulesen. Im anderen Unterordner namens „AufbereiteteDaten“ sind alle Werte zu finden, welche weiterverarbeitet und entsprechend selektiert wurden. Wie die Weiterverarbeitung erfolgte, ist im anschließenden Kapitel Datenvorverarbeitung zu lesen.

Dabei wurde auf zwei wesentliche Dateiformate für die Bereitstellung der Datei gesetzt. Diese sind CSV- und JSON-Dateien. Die CSV-Datei, auf welche der Scatterplot und die Parallelen Koordinaten zugreifen, heißt dabei „WineInformationExcelAufbereitetKlein“. Der Name der JSON-Datei, welche das Baumdiagramm verwendet, lautet „WineInformationGeoKleinKlein“. Dabei sind alle Daten innerhalb der CSV Datei mithilfe eines Kommas getrennt und das Dezimaltrennzeichen ist ein Punkt. Falls eine Null oder das Feld leer ist, so gilt dies als nicht vorhandener Wert. In der JSON-Datei hingegen wurden ausschließlich die Beziehungen zwischen den Weinen und den Ländern abgebildet. Somit haben die „data“-Felder nur eine „id“ zum Inhalt, welche den Namen enthält. Die Beziehung wurde mithilfe der „children“-Felder realisiert. Darüber hinaus wurden in der finalen JSON-Datei alle Länder entfernt, welche in der CSV-Datei keine Weine produzieren.

2.2 Datenvorverarbeitung

Bei der Datenverarbeitung wurden drei Schritten durchgeführt, um die Daten weiter zu bearbeiten. Diese Maßnahmen sind das Sichten-, Bearbeiten- und anschließende Überführung der Werte. Was in diesen Schritten genauer geschehen ist, wird nachfolgend erklärt.

Bei der Sichtung der Daten war es das Ziel, die Werte in ein Lesbares und leicht zu verarbeitendes Format zu bringen. Aufgrund dessen wurde das Dokument „cleasingWine“ in eine Exceldatei konvertiert, da eine Excel genau die Zielstellung dieses Schrittes erfüllt. Diese Datei ist unter dem Unterordner „AufbereiteteDaten“ mit dem Namen „WineInformationExcel“ zu finden. Durch diesen Schritt konnten diese Datenbestände nun einfacher eingelesen und bearbeitet werden. Bei der JSON-Datei wurde hierbei eine entsprechende zu ergänzende Datei gefunden, welche in der nächsten Maßnahme ergänzt werden konnte. Dieses Dokument wurde durch den GitHub Nutzer „Curran Kelleher“ zur Verfügung gestellt [6]. Die Datei ist dabei unter „AufbereiteteDaten“ mit dem Namen „WineInformationGeo“ zu finden und wurde mithilfe von „Visual Studio Code“ bearbeitet.

In der Bearbeitung der Werte sollten diese so aufbereitet werden, dass sie vom Code, welcher eingesetzt wurde, einfach zu verarbeiten sind. Dabei wurden sechs wesentliche Schritte getätigt, um diese Daten entsprechend bereitzustellen. Zuerst wurden aus den Spalten „sweet“, „acidity“, „body“ und „tannin“ die Namen, welche vor den Zahlen standen, entfernt, da es ansonsten nicht möglich gewesen wäre, diese in ein Zahlendatenformat zu konvertieren. Anschließend wurden die Zahlenwerte überarbeitet. So standen in der Spalte „abv“ und „degree“ eine Tilde, welche die minimale und maximalen Werte miteinander verbunden hat. Da jedoch kein Zahlendatenformat eine Tilde zulässt, wurden beide Zahlen getrennt und anschließend wurde der Durchschnitt gebildet. Falls nur ein Wert in der Spalte stand, wurde dieser übernommen. Darüber hinaus wurde der Betrag für die Weine in „price“ in südkoreanischen Won angegeben. Aufgrund dessen wurde dieser Preis mithilfe eines Währungskurses von 1 Euro zu 1.355,382 Won umgerechnet [15]. Nachdem diese Zahlen erfolgreich überarbeitet worden sind, wurden die Namen in der Spalte „name“ angepasst. So ist es bei der Konvertierung vorgekommen, dass Apostrophe und Umlaute nicht richtig importiert

worden sind. Diese Fehler wurden gesucht und entsprechend verbessert. Um die Daten anschließend besser in die JSON-Datei einfügen zu können, wurde außerdem eine neue Spalte erstellt, in welcher die Namen der Weine mit der entsprechenden JSON Schreibweise kombiniert wurden, um so das Zusammenfügen der Weinnamen und der Länder einfacher zu gestalten. Diese Änderungen wurden alle an der Exceldatei „WineInformationExcel“ durchgeführt. Anschließend wurde in einem neuen Dokument basierend auf „WineInformationExcel“, die Namen der Spalten vom Englischen ins Deutsche übersetzt. Zusätzlich dazu wurden „Column1“ und die Spalte mit den Schreibweisen für die JSON Datei herausgelöst. Dieser Stand findet sich in der Exceldatei mit dem Namen „WineInformationExcelAufbereitet“ wieder. Zum Abschluss dieser Bearbeitung der Daten wurden alle Datensätze herausgelöscht, welche in den Spalten „alc“ bis „ml“ einen nicht vorhandenen Wert enthielten. Diese Änderung ist in der Datei mit dem Namen „WineInformationExcelAufbereitetKlein“ wiederzufinden. Innerhalb der JSON-Datei gab es nur zwei Maßnahmen, welche zur Bearbeitung dieser vorgenommen worden sind. Im ersten Schritt wurden alle Weinnamen dort eingefügt, wo diese bereits auch in der CSV eine Zuordnung zu einem Land erhalten haben. Dieses Ergebnis ist unter der Datei „WineInformationGeoKlein“ zu finden. Anschließend wurden im letzten Schritt alle Länder entfernt, welche keine Weine nach der CSV-Datei hergestellt haben. Dies ist in der Datei „WineInformationGeoKleinKlein“ zu erkennen.

Die anschließende Überführung der Dokumente sollte die Werte wieder für den Programmcode lesbar machen. Somit wurden alle Informationen, welche in einer Exceldatei gespeichert waren, abermals in eine CSV-Datei zurücküberführt. Dabei wurde darauf geachtet, dass die Daten mithilfe eines Kommas getrennt worden sind und das Dezimaltrennzeichen ein Punkt ist. Die JSON-Datei wurde nicht umgewandelt, dementsprechend ist keine Überführung der Datei nötig.

Im Schritt „Bearbeitung der Daten“ wurden die verschiedenen Durchschnitte gebildet, um die Daten besser lesbar sowohl für die Menschen als auch die Maschine zu machen. Ansonsten hätten noch mehr Kategorien zur Auswahl gestanden, welche durch die kleine Differenz der Zahlen nicht sinnvoll gewesen wären. Weiterhin wurden verschiedene Datensätze in diesem Schritt entfernt. Dies wurde getan, um eine gemeinsame konsistente Datenbasis für alle Visualisierungen zu schaffen, sodass jeder Wein in allen Darstellungsweisen zu finden ist. Darüber hinaus führte die Reduktion des Datensatzes dazu, dass die Darstellungen nicht überfüllt wirken. Zusätzlich wurden alle Datensätze entfernt, welche nicht ins Deutsche übersetzt werden konnten.

3. Visualisierung

In den nachfolgenden Kapiteln wird genauer auf die einzelnen Visualisierungen und ihren Zweck eingegangen.

3.1 Analyse der Anwendungsaufgaben

Wie bereits im Kapitel Einleitung erwähnt worden ist, gibt es ein großes Hauptziel, welches mit diesen Visualisierungen erreicht werden soll. Dieses ist das die Zielgruppen ihre individuellen Fragen rund um das Thema Weine beantworten können. Dabei können vor allem Themata um die Eigenschaften des Weines beantwortet werden, welche innerhalb der ursprünglichen CSV-Datei vorhanden waren. Also Erkundigungen rund um Alkoholgehalt, Trinktemperatur, Süße, Säure, Körpers, Gerbstoffe, Preis, Jahres und Größe der Flasche des Weines. Diese Eigenschaften können nun mithilfe von den verschiedenen Visualisierungen gegenübergestellt werden und so die Fragen der unterschiedlichen Zielgruppen beantworten. Als Unterstützung für dieses Hauptziel gibt es noch drei etwas konkretere Fragestellungen, welche bereits in der Einleitung genannt worden sind. Auf diese wird nachfolgend eingegangen.

Um Fragen zwischen zwei Eigenschaften zu beantworten, sollte es eine Möglichkeit geben, zwei verschiedene Eigenschaften miteinander zu vergleichen. So könnten verschiedene Zusammenhänge zwischen den Weineigenschaften aufgedeckt und ein besseres Verständnis für diese geschaffen werden. Dadurch wird es möglich beispielsweise möglich einen preisgünstigen süßen Wein zu finden. Das Vorwissen der Zielgruppen sollte dabei so sein, dass diese wissen, was die Eigenschaften des Weines bedeuten, um korrekte Schlussfolgerungen aus den Darstellungen zu ziehen.

Um Fragen zwischen zwei Merkmalen zu beantworten, sollte es eine Möglichkeit geben, zwei unterschiedliche Eigenschaften miteinander zu vergleichen. So könnten verschiedene Zusammenhänge zwischen den Weineigenschaften aufgedeckt und ein besseres Verständnis für diese geschaffen werden. Dadurch wird es möglich, beispielsweise einen preisgünstigen süßen Wein zu finden. Das Vorwissen der Zielgruppen sollte dabei so sein, dass diese wissen, was die Eigenschaften des Weines bedeuten, um korrekte Schlussfolgerungen aus den Darstellungen zu ziehen.

In der Frage, ob es einen Zusammenhang zwischen den Produktionsmengen und der Anzahl der von Weinen in einer Nation, ist es nötig, zwei Informationen bereitzustellen. Dafür sind die Angaben der einzelnen Produktionsmengen der Länder nötig. Diese Information ist beispielsweise im „State of the Vitivinicultural World in 2020“ nachzulesen [11]. Unter dem Top 3 der am meisten produzierenden Nationen liegen Italien mit 49,1 mhl, Frankreich mit 46,6 mhl und Spanien mit 40,7 mhl. Diese Angaben sollten nun mit einer Darstellung verglichen werden, in welcher die Weine entsprechend ihren Ländern zugeordnet werden können, um so den Zusammenhang zwischen Produktionsmengen zu überprüfen.

Die eingesetzten Darstellungsformen des Scatterplots, Parallelen Koordinaten und Baumdiagramms helfen den Nutzenden, den vorliegenden Datensatz besser zu verstehen. So können diese mithilfe der Parallelen Koordinaten einen Überblick über den gesamten Daten erhalten und bereits gewisse Tendenzen oder Charakteristika erkennen. Falls dabei eine Besonderheit von zwei Eigenschaften oder eine genauere Analyse dieser nötig wird, können die Nutzenden dies anhand des Scatterplots durchführen. Weiterhin ist es möglich, mithilfe des Baumdiagramms herauszufinden, was die Weine für eine Herkunft besitzen.

Diese Struktur der Darstellungsformen ist dabei nicht zwingend notwendig, um die unterschiedlichen oben genannten Fragen beantworten zu können. Da diese verschiedenen Fragestellungen sich mit einer Darstellung beantworten lassen, ist eine Vernetzung der Abbildungen nicht nötig. Jedoch könnte eine solche Struktur erforderlich sein, um die Hauptfrage zu beantworten. Da hierbei ggf. je nach Frage verschiedene Darstellungsformen miteinander kombiniert werden müssen. Dann wäre eine solche Strukturierung sinnvoll und würde bei der Erkenntnisgewinn hilfreich sein.

3.2 Anforderungen an die Visualisierungen

Durch die Analyse der unterschiedlichen Ziele ergeben sich verschiedene Anforderungen an die Darstellungsformen. Für das Hauptziel ist es wichtig, dass die Visualisierungen so dargestellt sind, dass die Darstellungen entsprechend verständlich sind und trotzdem Fragen zu dem Thema Weine beantwortet werden können. Darüber hinaus sollten die verschiedenen Eigenschaften innerhalb der Anwendungen integriert sein.

Aus der Frage nach der Gegenüberstellung von zwei Merkmalen lassen sich zwei wesentliche Anforderungen an die Visualisierung ableiten. Dabei sollte es eine Auswahl geben, in welcher die Nutzenden nach ihren individuellen Wünschen Eigenschaften auswählen können, welche

anschließend automatisch angezeigt werden. Darüber hinaus sollten die Daten, welche dargestellt werden, so übersichtlich sein, dass sich Trends und Besonderheiten erkennen lassen.

Bei der Fragestellung, mehr als zwei Weineigenschaften zu vergleichen, sollte die Darstellung mehr als nur zwei Eigenschaften gegenüberstellen können. Dabei sollte jeder Datensatz trotzdem noch nachverfolgbar sein, um diesen gegeben zu identifizieren. Darüber hinaus sollte die Darstellung des Datensatzes so gut erfolgen, dass dieser leicht verständlich für alle Nutzenden ist und nicht beispielsweise zu überfüllt wirkt.

In der letzten Frage von Produktionsmengen und der Gesamtzahl von Weinen in einem Land sollte in der Darstellungsform erkennbar sein, wo diese herkommen. Darüber hinaus sollte die Summe der Weine in der Darstellung vermerkt sein. Weiterhin sollten die aktuellen Produktionsmengen der verschiedenen Nationen einsehbar sein, um diese mit der Anzahl der Weine im Land gegenüberzustellen. Zusätzlich wäre eine Darstellung, welche diese beiden Größen miteinander vergleicht, bei der Beantwortung dieser Frage sehr hilfreich.

3.3 Präsentation der Visualisierung

Nachfolgend werden die verschiedenen Visualisierungen vorgestellt, welche bei der Realisierung des Projektes verwendet worden sind. Dies sind der Scatterplot, die Parallelen Koordinaten und das Baumdiagramm.

3.3.1 Visualisierung Eins

Die erste Visualisierung innerhalb dieses Projektes ist ein Scatterplot, in welchem stets zwei verschiedene Eigenschaften des Weines gegenübergestellt werden. Dabei nimmt immer ein Merkmal eine Achse des Scatterplots ein und anschließend werden die Daten wie X- Y-Koordinaten in das entstandene Koordinatensystem eingetragen. Die Punkte werden in dieser Visualisierung als Kreise dargestellt. Wenn mit der Maus über diese gefahren wird, werden diese farblich markiert und der Text erscheint über diesem. Diese Beschriftung zeigt den Namen des Weins, das X und Y Merkmal dieses Datensatzes an. Darüber hinaus können die Eigenschaften angepasst werden. Dafür sind über dem Diagramm verschiedene Buttons zu finden, welche die Merkmale beinhalten. Wenn diese angeklickt werden, ändert sich je nachdem in welcher Zeile der Button ist, die jeweilige Achse des Scatterplots. Der Scatterplot ist in Abbildung 1 zu erkennen.

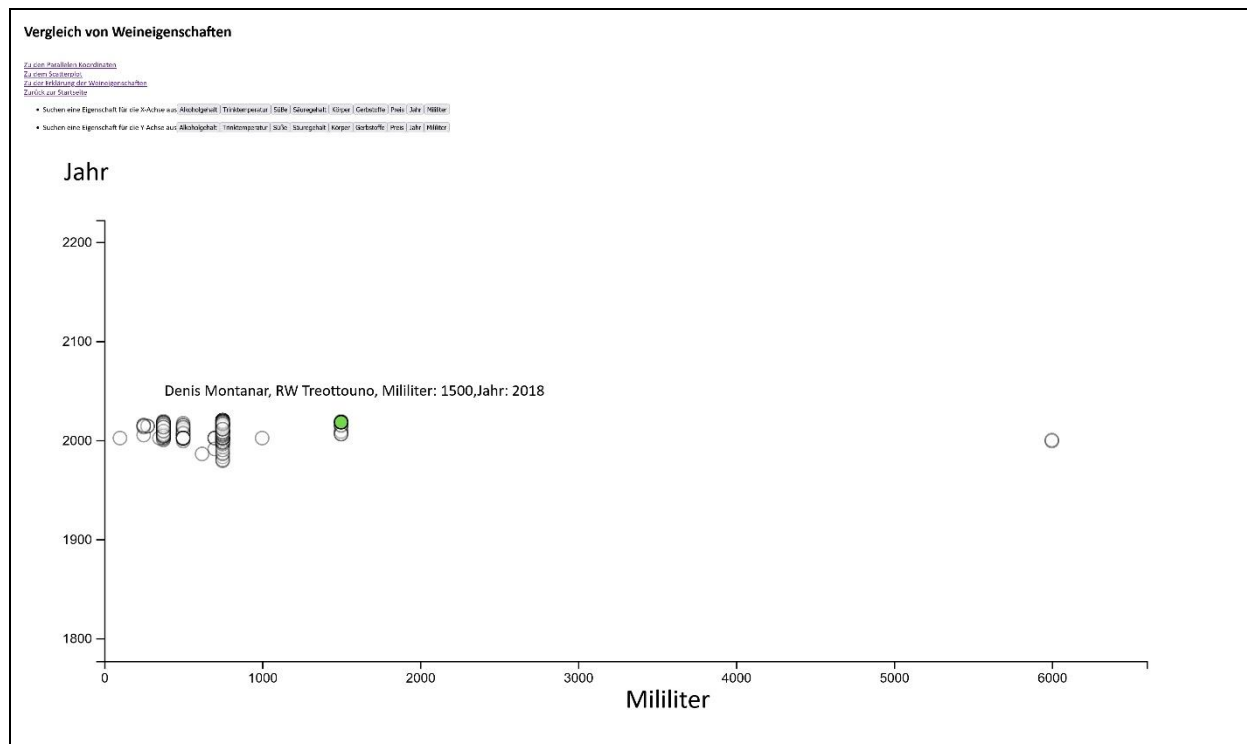


Abbildung 1: Scatterplot (Quelle: eigene Darstellung)

Die Anforderungen an den Scatterplot, welche es gibt, konnten erfüllt werden. So kann der Scatterplot zwei Merkmale des Weines gegenüberstellen. Drüber hinaus können die angezeigten Eigenschaften mithilfe der Buttons verändert werden. Diese Veränderungen passierten komplett ohne neu laden der Seite. Weiterhin lassen sich mithilfe der Darstellung der Koordinaten als Kreise entsprechende Trends oder Besonderheiten in den Daten gut erkennen.

In den Anforderungen dieses Projektes sollte die erste Darstellungsform ein Scatterplot oder ein Zeitreihendiagramm werden. Zeitreihendiagramme benötigen immer eine zeitliche Dimension in den Daten, um eine entsprechende Veränderung darstellen zu können. Diese Dimension gibt es jedoch nicht in dem vorliegenden Datensatz. Dementsprechend eignen sich diese Daten nicht, um über ein Zeitreihendiagramm dargestellt zu werden. Dadurch wurde bei der Darstellung auf den Scatterplot zurückgegriffen. Mithilfe von diesem Diagramm lassen sich zwei Eigenschaften gegenüberstellen und dies unabhängig von einer zeitlichen Dimension.

3.3.2 Visualisierung Zwei

Bei der zweiten Visualisierung handelt es sich um ein Paralleles Koordinaten Diagramm. In einer solchen Darstellungsform werden die Eigenschaften der Daten als Achse dargestellt. Diese werden durch Linien miteinander verbunden. Eine Linie stellt dabei einen Datensatz dar und dessen Werte auf den unterschiedlichen Achsen. So gibt es in dieser Darstellung vier verschiedene Achsen, welche die Eigenschaften der Weine darstellen können. Die dargestellten Linien sind dabei die unterschiedlichen Weine. Die Achsen können dabei individuell den Eigenschaften zugewiesen werden. Dies erfolgt mithilfe von verschiedenen Buttons, welche über der Darstellung zu finden sind. Wenn dabei ein Button gedrückt wird, nimmt die entsprechende Achse das gewünschte Merkmal an. Zusätzlich werden, wenn mit der Maus über eine Linie gefahren wird, die Eigenschaften angezeigt und die Linie farblich hervorgehoben. Die Parallelen Koordinaten sind in Abbildung 2 zu erkennen.

Vergleich von mehrern Weineigenschaften

- [Zur Beschreibung](#)
[Zu einer Suchansicht](#)
[Zurück zur Startseite](#)
- Suchen eine Eigenschaft für die erste Spalte aus: Alkoholgehalt | Trinktemperatur | Süße | Säuregehalt | Körper | Gerbstoffe | Preis | Jahr | Mitter
 - Suchen eine Eigenschaft für die zweite Spalte aus: Alkoholgehalt | Trinktemperatur | Süße | Säuregehalt | Körper | Gerbstoffe | Preis | Jahr | Mitter
 - Suchen eine Eigenschaft für die dritte Spalte aus: Alkoholgehalt | Trinktemperatur | Süße | Säuregehalt | Körper | Gerbstoffe | Preis | Jahr | Mitter
 - Suchen eine Eigenschaft für die vierte Spalte aus: Alkoholgehalt | Trinktemperatur | Süße | Säuregehalt | Körper | Gerbstoffe | Preis | Jahr | Mitter

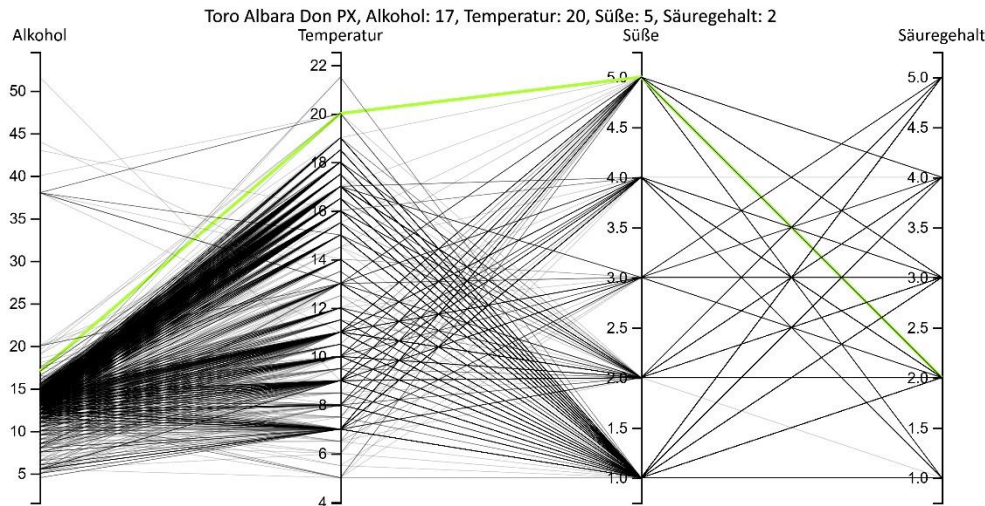


Abbildung 2: Parallele Koordinaten (Quelle: eigene Darstellung)

Die Anforderungen an diese Darstellung konnten erfüllt werden. So ist es mithilfe dieser Darstellungsform möglich, die verschiedenen Eigenschaften gegenüberzustellen und individuell anzupassen. Weiterhin werden Werte angezeigt, wenn innerhalb dieses Diagramms über eine Linie drübergefahren wird. Somit ist die Verfolgbarkeit und Nachvollziehbarkeit der Daten gegeben. Durch das Erfüllen der Übersichtlichkeit und der Nachverfolgung der Werte sind die Anforderungen an dieses Diagramm erfüllt.

Innerhalb der Anforderungen für das Projekt sollte die zweite Darstellung eine mehrdimensionale Darstellungsform sein. Für diese würden die Scatterplots, Projektion und Selektion, Parallelen Koordinaten, K-Means und Datentinte infrage kommen. Bei der Datentinte wird versucht, sich auf die Hauptaussage der Darstellung zu konzentrieren und somit den Teil der Visualisierung zu löschen, welcher verlustfrei entfernt werden kann. Da jedoch aufgrund der verschiedenen Zielgruppen (vor allem der Weineinkäufer und Weinexperten) keine Werte verloren gehen sollen, fällt diese für die Darstellungsform von mehrdimensionalen Daten heraus. Dasselbe gilt für die K-Means, welche versuchen den Abstand (quadriert) der Punkte zu den Prototypen zu minimieren. Dies würde mit einem Datenverlust einhergehen, was durch die Zielgruppen nicht erwünscht ist. Bei der Projektion und Selektion wird versucht, ein mehrdimensionalen Raum zu visualisieren. Dafür wird dieser mithilfe von Projektionen auf verschiedenen 2D Darstellungen abgebildet. Da diese Darstellungsform jedoch viele Visualisierungen nach sich ziehen würde, wäre dies nicht im Interesse der Übersichtlichkeit und schließt sich aus den Anforderungen an dieses Diagramm aus. Gleichmaßen müssten bei den Scatterplots die mehrdimensionalen Daten auf verschiedene Scatterplots aufgeteilt werden und widersprechen also auch den Forderungen. Die Parallelen Koordinaten können mehrere Dimensionen in einer Darstellung darstellen und somit trotzdem eine gewisse Übersichtlichkeit garantieren. Deswegen wurde die zweite Visualisierung mithilfe der Parallelen Koordinaten umgesetzt.

3.3.3 Visualisierung Drei

Bei der letzten und somit dritten Visualisierung handelt es sich um Baumdiagramm. Mithilfe eines solchen Diagramms können verschiedene hierarchische Beziehungen dargestellt werden. So sind diese untereinander angeordnet und anhand von Linien verbunden. In der aktuellen Darstellung werden die Länder und Weinnamen mithilfe von Kreisen dargestellt. Diese werden anschließend je nach Beziehung mit einer Linie verbunden. Der Ausgangskreis ist „World“. Bei der Baumhierarchie gibt es keine weiteren Interaktionsmöglichkeiten. Das Baumdiagramm ist in Abbildung 3 zu finden.

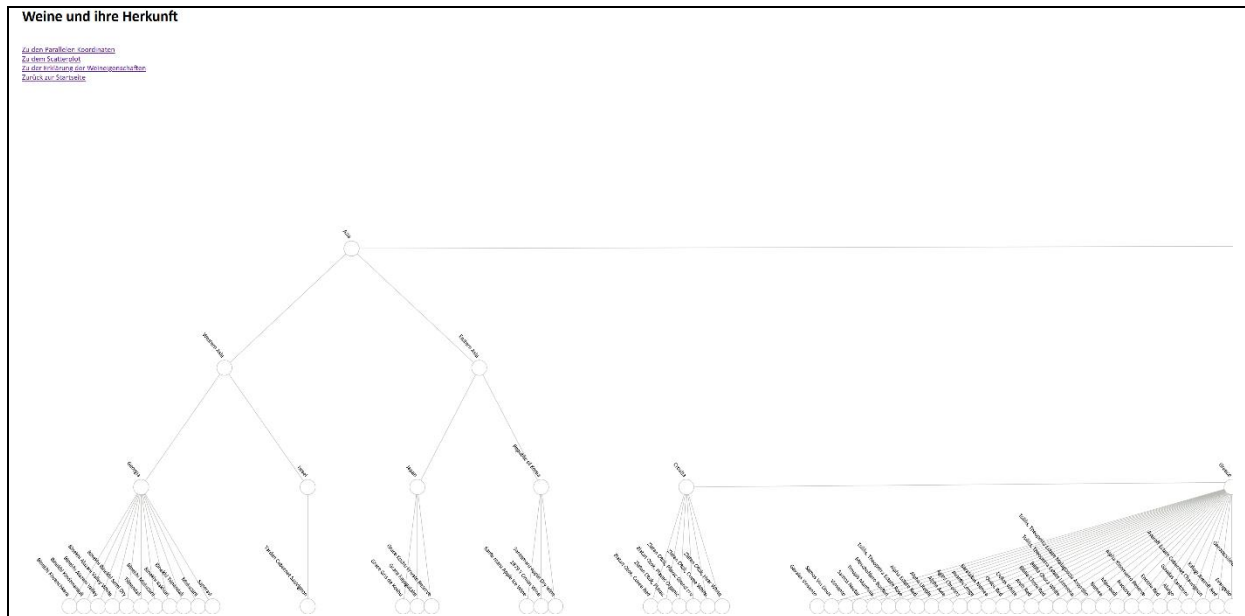


Abbildung 3: Auszug aus der Baumhierarchie

Die Anforderungen an das Baumdiagramm wurden nur teilweise erfüllt. Dabei sollte mithilfe dieser Darstellung die Frage nach den Produktionsmengen und Weinarten beantwortet werden. So stellt diese Hierarchie zwar die Länder und deren Weine dar, jedoch gibt es keinen Aufschluss darüber, wie viele diese sind. Weiterhin ist kein Abgleich der Produktionsmengen der einzelnen Nationen möglich, da diese nicht aufgeführt worden sind. Dementsprechend gibt es auch keine Gegenüberstellung dieser beiden Größen. Somit wurden die Anforderungen nur teilweise erfüllt, da die Länder und ihre Weine zwar dargestellt worden sind, jedoch fehlen alle anderen Ansprüche.

Für diese Darstellungsform sollte nach den Anforderungen des Projektes eine Baum- oder Graphen-Technik verwendet werden. Beide Techniken eignen sich Zusammenhänge zu behandeln und somit für die Visualisierung von Hierarchien. In dieser Darstellung mussten jedoch nur schlichte hierarchische Verbindungen angezeigt werden, da durch die Zusammenhänge von Ländern und Weinen eine mehrfache Verweisung auf einen Punkt unnötig ist. Dementsprechend wurde sich aufgrund der Übersichtlichkeit für ein Baumdiagramm entschieden, bei welchem es verschiedene Techniken gibt, solche einfachen hierarchischen Beziehungen übersichtlich darzustellen.

3.4 Interaktion

Bei den unterschiedlichen Darstellungen gibt es verschiedene Interaktionsmöglichkeiten, welche betätigt werden können. Die Hauptinteraktionsart ist dabei der Button. Mithilfe dieser können die bei den einzelnen Darstellungsformen die Eigenschaften, welche aktuell angezeigt werden, geändert werden. Damit ist es dem nutzenden Möglichen, die Darstellungen so anzupassen, wie diese es benötigen. Weiterhin besitzen diese die Möglichkeit mithilfe des Schwebens der Maus über die Kreise des Scatterplots oder den Linien der Parallelen Koordinaten den genauen Datensatz angezeigt

bekommen. Damit können die Nutzenden die verschiedenen Daten identifizieren und so die besonderen Werte der entsprechenden Darstellung herausfinden. Die Buttons sollten dabei jedem Nutzenden bekannt sein, weswegen auf diese Darstellungsform zurückgegriffen worden ist. Zusätzlich können die Nutzer die Daten des Scatterplots und der Parallelen Koordinaten genauer untersuchen, indem diese mithilfe der Maus über die Kreise oder Linien dieser schweben. Auf diese Darstellungsform wurde zurückgegriffen, da diese am übersichtlichen ist und somit auch nicht die Erkennung von verschiedenen Trends oder Besonderheiten beeinflusst. Weitere Interaktionsmöglichkeiten wurden aufgrund der Übersichtlichkeit und Einfachheit nicht verwendet.

4. Implementierung

Das Projekt wurde auf Basis der verschiedenen Übungen angefertigt, welche bereits vor der Aufgabe fertiggestellt wurden. Dabei dienten insbesondere die erste und dritte Übung mit dem Scatterplot und der Interaktionsmöglichkeit als Grundlage für den Scatterplot. Bei den Parallelen Koordinaten wurde die Übung sieben als Ausgangspunkt genutzt. Innerhalb der Baumhierarchie wurde die zehnte Übung als Fundament benutzt. Zusätzlich wurde für den Scatterplot und die Parallelen Koordinaten die Übung acht mit ihrem CSV-Decoder als Ausgangspunkt eingesetzt. Dabei wurde bei beiden Darstellungen jeweils die Übung als Grundlage verwendet und versucht diese mit dem CSV-Decoder zu verbinden, um so das Importieren der CSV-Datei zu ermöglichen. Nachdem dies erfolgreich umgesetzt worden ist, wurde für den Scatterplot noch die Buttons eingebaut, um diesen entsprechend interaktiv zu machen. Dafür waren weitere Anpassungen am CSV-Decoder und am Code nötig, welche auch anschließend bei der Entwicklung der parallelen Koordinaten einfließen könnten. Bei der Baumhierarchie wurde musste wenig angepasst werden, da bereits in der Übung ein JSON-Decoder enthalten war. Hierbei war es nur wichtig, die entsprechenden Daten bereitzustellen. Zuletzt wurden die Visualisierungen noch einmal in ihrer Darstellungsform überarbeitet. Die Entwicklung dieser Darstellungen erfolgte dabei im Wesentlichen nacheinander, wobei sich einige Weiterentwicklungen überlagerten, um die Zeit der Fehlersuche zu überbrücken. So wurde zuerst der Scatterplot entwickelt, anschließend die Parallelen Koordinaten und zuletzt die Baumhierarchie.

Der Quellcode ist dabei in verschiedene Parts aufgeteilt und strukturiert. Diese Struktur wurde bei allen Visualisierungen angewendet, um ein schnelleres Zurechtfinden zu garantieren. Zuerst stehen immer die zu importierenden Elm-Pakete, welche für diesen Abschnitt benötigt werden. In allen drei verschiedenen Darstellungen wurde dabei auf die Packages „Html“ und „TypedSVG“ zugegriffen. Anschließend sind alle „type“ und „type alias“ definiert, welche innerhalb des Quellcodes vorhanden sind. Danach finden sich weitere Funktionen, welche für die verschiedenen Parts von Elm brauchen. Begonnen wird dabei immer mit den Decodern der Dateien und anschließend werden alle Funktionen angeführt, welche für den Part benötigt werden, zuletzt werden die Elm Parts „main“, „init“, „subscriptions“, „view“ und „update“ aufgeführt. Die Daten werden innerhalb eines Records abgespeichert, welcher innerhalb der Main zu finden ist. Auf diesen Record wird anschließend zugegriffen, wenn eine Aktualisierung stattfindet.

Ein besonderes Problem der Darstellungen war, den CSV-Decoder zu erweitern und mit den Übungen zu verbinden. Bei der Erweiterung des CSV-Decoders lag die Herausforderung darin, mehr als 4 Spalten auf einmal zu verarbeiten, da Tupels von Elm nur maximal drei verschiedene Items akzeptieren. Dies wurde mithilfe eines neu geschriebenen Decoders umgangen. Bei der Verbindung des CSV-Decoders und der Übung gab es primär Probleme mit den Funktionen und den Datentypen, welche aber mit verschiedenen Anpassungen der Funktionen gelöst werden konnten. Die eigentliche Darstellung wurde jedoch annähernd so wie aus der Übung adaptiert und nahm somit

verhältnismäßig wenig Zeit in Anspruch. Dasselbe gilt für die Baumhierarchie, welche fast ohne größere Anpassungen der Übung übernommen werden konnte.

5. Anwendungsfälle

Nachfolgend wird für die drei Visualisierungen jeweils ein Anwendungsfall vorgestellt. In diesem wird die entsprechende Darstellung verwendet, um verschiedene Informationen aus dieser zu gewinnen.

5.1 Anwendung Visualisierung Eins

In dem ersten Anwendungsfall wird der Preis mit dem Körper eines Weines verglichen. Dies erfolgt mithilfe eines Scatterplots und ist in Abbildung 4 zu erkennen.

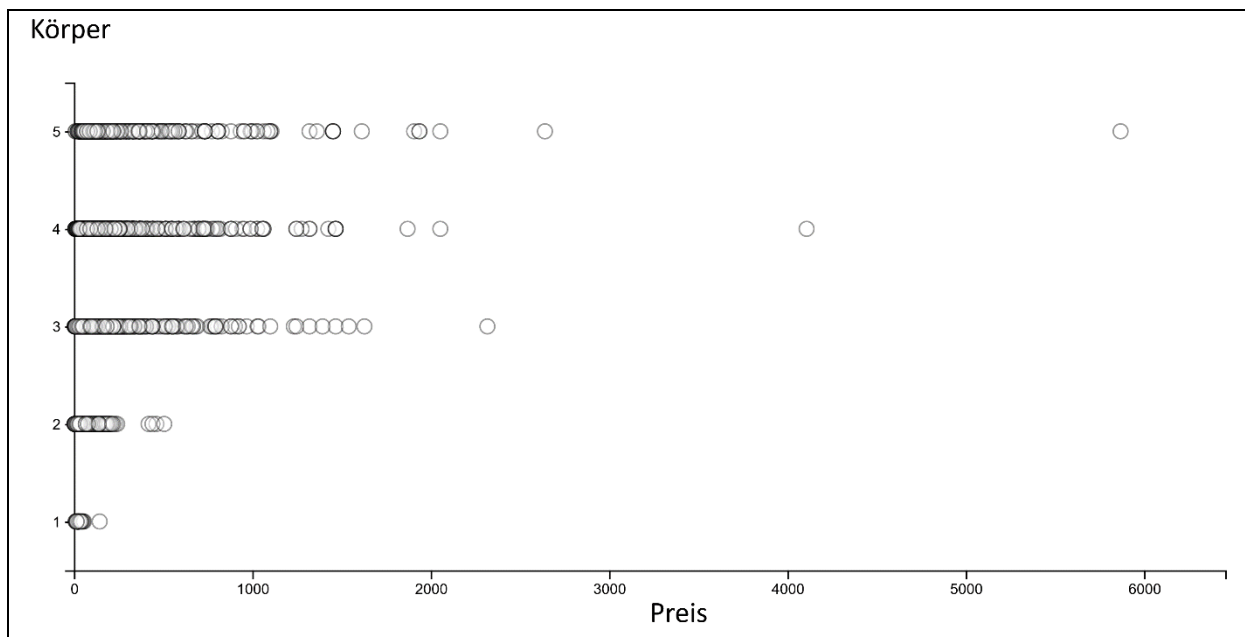


Abbildung 4: Scatterplot welcher die Eigenschaften Preis und Körper gegenüberstellt (Quelle: eigene Darstellung)

In Abbildung 4 ist zu entnehmen, dass es selbst für einen niedrigen Betrag Weine in allen Körpergrößen gibt. Jedoch ist klar zu erkennen, dass mit steigender Körpergröße der Preis ebenso immer weiter ansteigen kann. So ist besitzt der teuerste Wein auch die größte Körperform. Weiterhin wächst die Verteilung der Weine auf der Preisachse mit steigender Körpergröße an, bis dieser bei einer Körperform von 3 bis 5 gleichbleibend ist. Weswegen davon auszugehen ist, dass mit ansteigendem Preis auch die Körpergröße steigen sollte.

Diese Darstellung könnte so von einem Weininteressierten verwendet werden, um herauszufinden, wie er unter einem guten Preis-Leistungs-Verhältnis einen Wein mit einem großen Körper erhält. Hierbei sollte dieser Zielgruppe auffallen, dass er verschiedenen Körpergrößen auch zu einem geringen Preis bekommt. Jedoch je mehr diese Gruppe bereit ist, an Geld für den Wein auszugeben, desto wahrscheinlicher ist es, dass diese einen Wein mit einem größeren Körper erhält.

Als alternative zu diesen Diagrammen hätte es ein Zeitreihendiagramm in diesem Projekt visualisiert werden können. Da jedoch bei diesen Daten keinen zeitlichen Hintergrund gibt, ist es nicht möglich, dieses Problem über Zeitreihendiagramme darzustellen.

5.2 Anwendung Visualisierung Zwei

Bei diesem Anwendungsfall werden mithilfe der parallelen Koordinaten die verschiedenen Eigenschaften des Körpers, Gerbstoffe, Süße und Säuregehalt des Weines gegenübergestellt. Diese Darstellung ist dabei in Abbildung 5 zu erkennen.

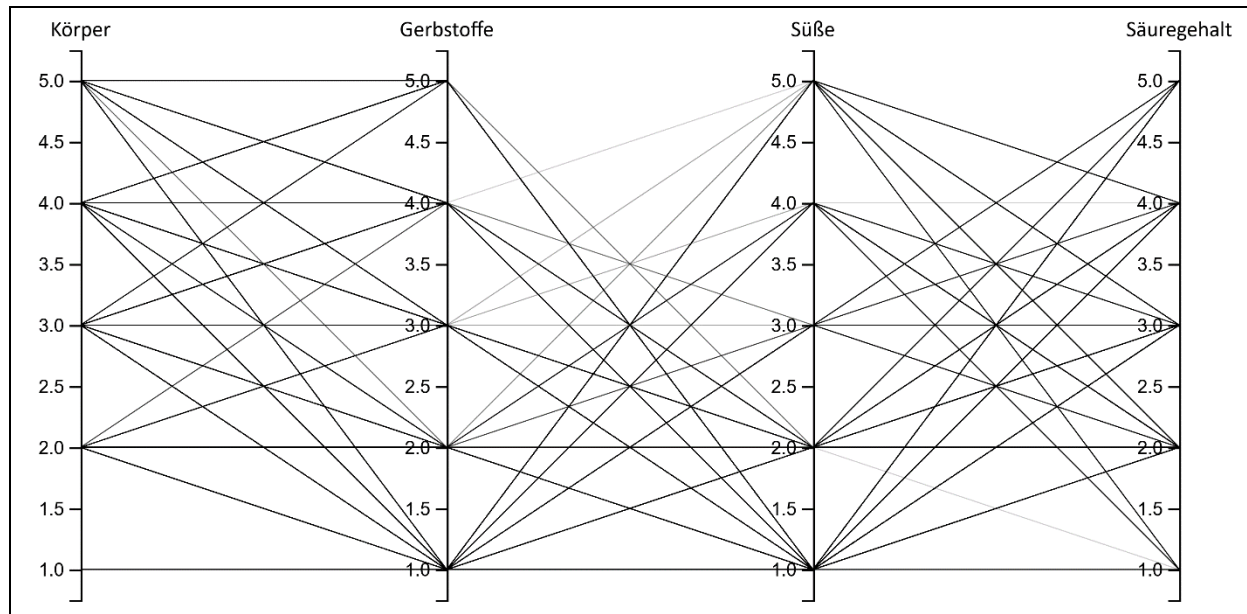


Abbildung 5: Gegenüberstellung der Eigenschaften Körper, Gerbstoffe, Süße und Säuregehalt mithilfe der Parallelen Koordinaten (Quelle: eigene Darstellung)

Dabei ist in Abbildung 5 zu sehen, dass es keine besonderen Muster zwischen den verschiedenartigen Merkmalen auftreten. Dementsprechend sind verschiedene Möglichkeiten vorhanden, die unterschiedlichen Eigenschaften in abweichender Intensität in einem Wein wiederzufinden. Auffälligkeiten an den Daten gibt es allerdings schon. Somit kann ein Wein, welcher die maximale Süße von fünf hat, nur einen Säuregehalt von vier aufweisen. Umgekehrt ist es ähnlich, so kann ein Wein mit einem Säuregehalt von fünf maximal eine Süße von drei besitzen. Somit kann es keinen Wein geben, welcher auf beiden Skalen eine Fünf besitzt. Dies hängt mit dem Gärungsprozess der Weintraube zusammen, was bereits im Kapitel Das Hauptziel ist dabei, diesen Datensatz so bereitzustellen, dass die unterschiedlichen Zielgruppen ihre individuellen Fragen rund um das Thema Weine beantworten können. Um dies zu erfüllen, sollten zwei andersartige Charakteristika des Datensatzes gegenübergestellt werden können. Somit wäre es beispielsweise denkbar, das PreisLeistungsverhältnis bei den unterschiedlichen Merkmalen herauszufinden. Weiterhin sollte es möglich sein, mehr als nur zwei Charakteristika Datensatzes zu vergleichen, um so ein besseres Verständnis für die Eigenschaften und Zusammenhänge des Weines zu schaffen. Außerdem sollten die Visualisierungen eine Möglichkeit der Zuordnung der Weine nach der Herkunft ermöglichen, umso beispielsweise überprüfen zu können, ob die verschiedenen Produktionsmengen mit den Weinarten eines Landes gleichmäßig mitwachsen. Diese Ziele werden anschließend in diesem Projektbericht mithilfe von drei Visualisierungstechniken bearbeitet.

Anwendungshintergrund erklärt worden ist. Eine weitere Besonderheit besteht zwischen der Süße und den Gerbstoffen. So kann ein Wein, welcher einen hohen Gerbstoffgehalt von fünf besitzt, maximal eine süße von zwei erreichen. Dies setzt sich aber nicht in den unteren Stufen fort. Dort ist es bereits ab einem Gerbstoffgehalt von vier möglich, die volle Süße von fünf anzutreffen. Jedoch gibt es dementsprechend auch wie bei der Säure und der Süße keinen Wein, welcher eine Fünf in den Gerbstoffen und in der Süße erreichen kann. Umgekehrt verhält es sich bei der Beziehung zwischen den Körper und den Gerbstoffen. Hier besitzt ein Wein mit der Körpergröße eins auch nur

einen Gerbstoffgehalt von eins. Jedoch besitzen die restlichen Körpergrößen alle verschiedene Gerbstoffgehalte.

Dieser Anwendungsfall wäre etwas für die Zielgruppe der Weininteressierten. Dabei könnten diese versuchen herauszufinden, welche Zusammenhänge es zwischen den unterschiedlichen Geschmacksempfindungen auftreten. Dabei sollte dieser Zielgruppe auffallen, dass es pauschal bis auf wenige Ausnahmen viele Möglichkeiten gibt, die verschiedenen Geschmäcker miteinander zu kombinieren. Dabei können jedoch beispielsweise Einschränkungen auftreten, dass es eine Süße von fünf und einen Säuregehalt von fünf nicht möglich ist.

Bei dieser Visualisierung gäbe es verschiedene Alternativen zu dieser Diagrammart. Diese wären Scatterplots, Projektion und Selektion, K-Means und Datentinte. Da die K-Means und Datentinte Darstellungen waren, in denen etwas weggelassen worden ist, fallen diese heraus, da der Weininteressierte alle Weine miteinander vergleichen möchte. Bei den Scatterplots, Projektion und Selektion würde es unterschiedliche Grafiken geben, in denen die verschiedenen Merkmale gegenübergestellt werden. Dies würde zwar zu einer besseren Analyse der Daten führen, da jede Eigenschaft direkt nebeneinandergehalten werden kann, was aktuell bei den parallelen Koordinaten nicht geschieht. Jedoch würde dies gleichzeitig auch zu einem gewissen Informationsverlust einhergehen. Dieser würde daher kommen, dass die verschiedenen Eigenschaften einzeln gegenübergestellt werden würden und somit Trends oder Muster nicht so einfach zu erkennen sind.

5.3 Anwendung Visualisierung Drei

Innerhalb des letzten Anwendungsfalls werden die verschiedenen Weine von Asien genauer analysiert. Ein Ausschnitt aus dem Baumdiagramm, welche die Weine in Asien darstellt, ist in Abbildung 6 zu erkennen.

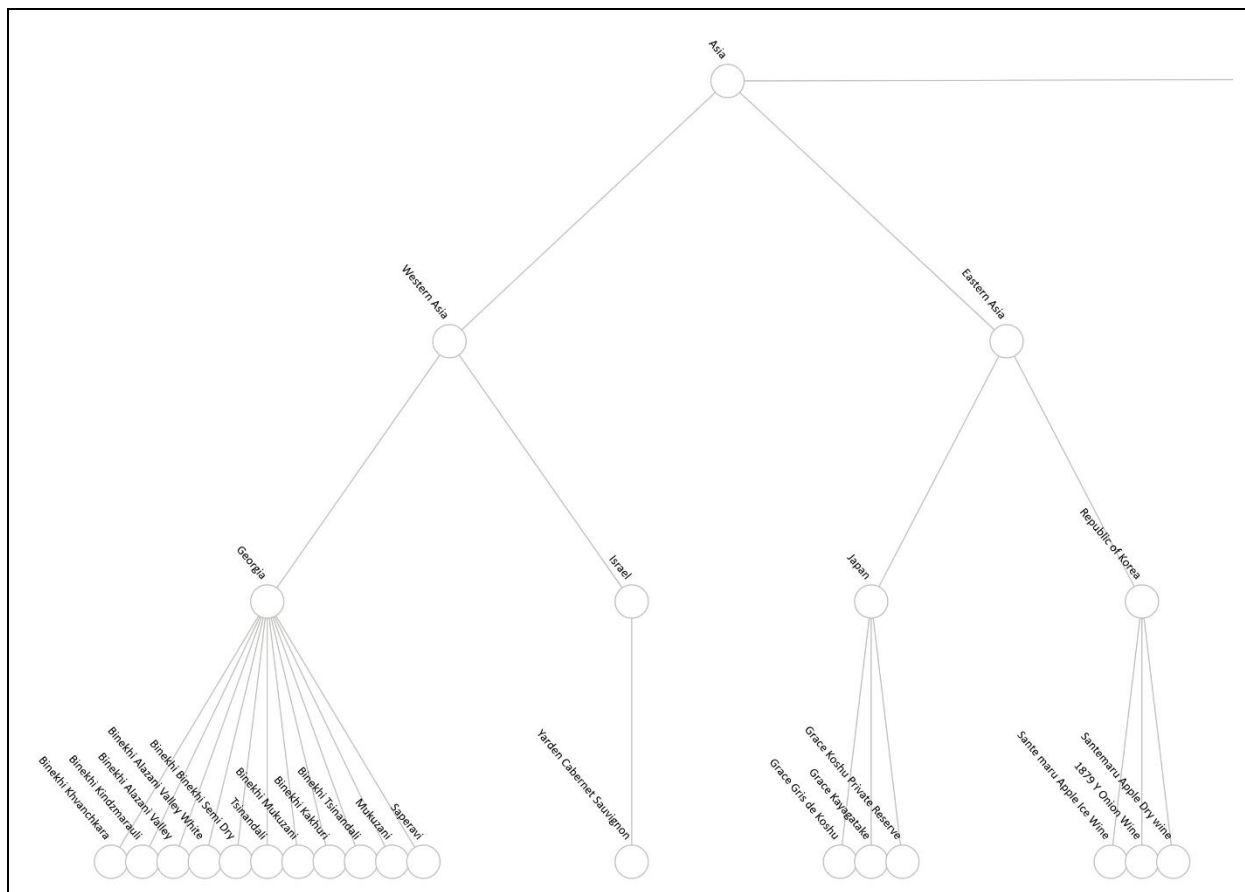


Abbildung 6: Ausschnitt mit Asiatischen Weinen aus der Baumhierarchie (Quelle: eigene Darstellung)

In Abbildung 6 sind die verschiedenen Weinamen zu sehen, welcher der Datensatz an asiatischen Weinen zu bieten hat. Dabei ist auffällig, dass Georgien die meisten Weinarten mit elf Stück besitzt. Die niedrigste Gesamtmenge hat Israel mit nur einer Weinart. Die beiden anderen Nationen Japan und Südkorea sind mit einer Weinarten Anzahl von drei vorhanden. Auffällig ist dabei, dass es trotz der vielen Staaten in Asien nur vier Länder in diesem Datensatz vertreten sind, welche Weine besitzen.

Diese Visualisierung könnte von einem Weineinkäufer verwendet werden, welcher gerne auf dem heimischen Markt asiatische Weine verkaufen möchte. So könnte er mithilfe dieser Darstellung herausfinden, welche verschiedenen Weine es aus dem asiatischen Raum gibt und aus welchem Land diese kommen. Damit könnte dieser die Transportkosten oder eine weitere Spezialisierung seines Geschäfts vorzunehmen.

Als Alternative für diese Darstellung gab es die Möglichkeit, die verschiedenen Weinarten mithilfe der Graphen-Technik zu visualisieren. Da es sich bei diesen Daten um eine einfache Hierarchiebeziehung handelt und somit nur eine Beziehung zwischen den einzelnen Kreisen dargestellt werden muss, ist diese Graphen Technik überflüssig.

6. Verwandte Arbeiten

Die hierbei verwandten Arbeiten beschäftigen sich mit der Visualisierung von medizinischen Daten. Dies wird dabei bei beiden Dokumenten unter anderem mit einem Scatterplot realisiert. Nachfolgend werden die beiden Arbeiten vorgestellt und Unterschiede zwischen diesem und der vorgestellten Text hervorgehoben.

Beim ersten Dokument handelt es sich um die Visualisierung von Patientendaten, welche an Multiplen Sklerose leiden. Ziel dieser Arbeit war es dabei, die anfallenden Daten dabei mithilfe eines Scatterplots darzustellen. Die daraus resultierende Benutzeroberfläche sollte so einfach wie möglich sein und Interaktionen der Nutzenden zulassen. Um dies zu verwirklichen, wurden die Werte bereits im vornherein mithilfe von unterschiedlichen Klassen eingeteilt, was der Übersichtlichkeit der Daten dienen sollte. Zusätzlich dazu sollte es in der Darstellung möglich sein, verschiedene statistischen Größen wie beispielsweise den Median sich zu berechnen lassen, was entsprechend realisiert werden musste. Die Daten werden dabei mithilfe der Middleware „ColdFusion“ aufbereitet. Weiterhin werden die verschiedenen Informationen mithilfe dieser Software dargestellt. Dabei wurden in dieser Arbeit ein Scatterplot, Balkendiagramm und ein Histogramm erstellt. Eine Interaktionsmöglichkeit gab es dabei ausschließlich mit dem Balkendiagramm. Dabei war es möglich, über ein Dropdownmenü sich alle Eigenschaften der Datenbank anzuzeigen lassen und eine entsprechende Auswahl zu treffen. Dies ist in der Darstellung sowohl für die X- als auch die y-Achse durchführbar. Darüber hinaus ist es möglich, nach anschließender Auswahl sich ein Histogramm zu erzeugen lassen. Bei dem Scatterplot wurden darüber hinaus noch die Konfidenzintervalle und eine Regressionsgerade hinzugefügt. [10]

Die Gemeinsamkeiten zu dieser Arbeit liegen dabei in der Zielstellung, so sollte ebenfalls in diesem Projekt eine Benutzeroberfläche erstellt werden, welche Interaktionen mit den Nutzenden zulässt. Darüber hinaus wurde sich auch für eine Darstellung der Daten als Scatterplot entschieden. Jedoch sind die Interaktionsmöglichkeiten bei dieser Arbeit weitreichender als bei der hier vorliegenden. So ist es möglich, mithilfe der Interaktionsmöglichkeit bei den Balkendiagrammen sich ein Histogramm zu erstellen. Anstatt der Buttons, welcher in dieser Arbeit verwendet wurden, wurde auf ein

Dropdownmenü gesetzt, um die Auswahl der verschiedenen Achsen zu realisieren. Zusätzlich wurden beide Projekte in unterschiedlichen Programmiersprachen umgesetzt. [10]

Im zweiten verwandten Text geht es um die Visualisierung von verschiedenen Daten rund um Diabetespatienten. Dabei fokussierte sich diese Arbeit besonders auf die Analyse der Nutzbarkeit des Scatterplots. So sollte untersucht werden, in wie mit diesem Scatterplot umgegangen wird. Der Scatterplot besitzt dabei unterschiedliche Interaktionsmöglichkeiten. Dementsprechend ist es realisierbar, die verschiedenen Achsen mithilfe von einem Dropdownmenü anzupassen und die Daten können nach dem Wünschen der Nutzenden angepasst werden. Darüber hinaus ist es machbar, auf die unterschiedlichen Werte herein- und herauszoomen. Des Weiteren ist es möglich, Hilfe des Berührens des Punktes mit der Maus mehr Informationen zu der entsprechenden Untersuchung angezeigt zu bekommen. Da es sich dabei um Werte handelt, welche eine zeitliche Dimension besitzen, ist es in diesem Scatterplot praktikabel, mithilfe der Bedienung eines Videorekorders sich zwischen den verschiedenen Zeitpunkten hin und her zu bewegen. Um dies zu erleichtern, gibt es weiterhin eine Funktion, um die Daten anhand einer Spur verfolgen zu können.[8]

Der große Unterschied zwischen dieser und der eben vorgestellten Arbeit ist, dass der Scatterplot um viele Funktionen erweitert wurde, um die Bedienung für die Nutzenden zu vereinfachen. So gibt es in beiden die Möglichkeit, die Achsen durch Buttons oder Dropdown Menüs zu verändern. Weiterhin ist es möglich, in beiden Informationen über den Datensatz zu erhalten, indem der Punkt mithilfe der Maus berührt wird. Jedoch gibt es durch die zeitliche Dimension viel mehr Möglichkeiten, diese Daten anders im Scatterplot darzustellen und somit mehr Interaktionsmöglichkeiten zu verwenden. Dies wurde zum Beispiel durch die Art Videofernbedienung umgesetzt. Aber auch andere Funktionen fehlen in dieser Arbeit, welche nicht von der zeitlichen Dimension abhängig sind. Hier wäre beispielsweise das Zoomen auf eine neue Ebene, um die Daten besser erkennen zu nennen. Zusammenfassend ist zu sagen, dass dieses Dokument mehr Darstellungen als die vorgestellte Arbeit besitzt. Jedoch hat diese viel mehr Interaktionsmöglichkeiten, um die Nutzungserfahrung so einfach und angenehm zu gestalten wie nur möglich. [8]

7. Zusammenfassung und Ausblick

Mithilfe dieser Projektarbeit ist es möglich, den Weindatensatz, welcher die Grundlage für dieses Projekt bildet, einfach und mithilfe verschiedenster Darstellungsformen zu analysieren. Diese Darstellungen besitzen dabei eine unkomplizierte Benutzeroberfläche, um mit diesen zu interagieren und es auf die entsprechenden Wünsche der Nutzenden anzupassen. Dadurch ist eine einfache und individuell angepasste Analyse der Daten realisierbar. Somit ist es den verschiedenen Nutzenden möglich, verschieden Trends, Muster oder auch unterschiedliche Zusammenhänge in diesen Werten zu finden, nachzuvollziehen oder zu reproduzieren.

Die Zielgruppen, für welche diese Arbeit geschrieben worden ist, erhalten durch diese Visualisierungen diverse Vorteile. So ist es den Weininteressierten nun möglich, durch den Scatterplot unterschiedliche Weine zu finden, welche seinem Geschmack entsprechen oder komplett neue Sorten zu entdecken. Darüber hinaus ist es beispielsweise machbar, mithilfe der Parallelen Koordinaten und der Baumhierarchie verschiedene Zusammenhänge einfacher zu verstehen. Auch die Zielgruppe der Weinexperten erhält verschieden Vorteile. Diese können mithilfe der unterschiedlichen Visualisierungen neue Sorten entdecken oder ihre aktuellen Weine einfacher einschätzen, um so ein noch besseres Wissen an Weinen anzusammeln. Weiterhin sollte es möglich sein, mithilfe dieser verschiedenen Darstellungen ihr Weinwissen problemloser darzustellen und nachzuvollziehen. Innerhalb der Gruppe der Weinverkäufer ist es realisierbar, neue Weine für die Kunden zu finden und diesen damit eine noch bessere Beratung zu geben. Darüber hinaus ist es

möglich, mithilfe des Baumdiagrammes eine Spezialisierung auf einen gewissen Kontinent oder Land umsetzen.

Eine Erweiterung dieses Projektes bei den Visualisierungen wäre vor allem bei der Interaktivität der einzelnen Darstellungen denkbar. So wäre es möglich, eine Filterfunktion für die Daten einzubauen, oder auch eine Hervorhebung der Werte, welche aktuell ausgewählt sind. Zusätzlich könnten weitere Darstellungen hinzugefügt werden, um eine Analyse der verschiedenen Informationen noch einfacher zu gestalten. So wäre beispielsweise ein Balkendiagramm denkbar, in dem die einzelnen Eigenschaften eines Weines aufgeführt werden. Weiterhin sollte bei steigender Anzahl der Datenmengen über eine Kategorisierung der Daten nachgedacht werden, um somit die Übersichtlichkeit innerhalb der Darstellungen nicht zu verlieren. Bei der Baumhierarchie wäre es denkbar, die einzelnen Bäume erst per Mausklick zu öffnen, umso eine noch bessere Sichtbarkeit zu gewährleisten.

Bei der Datenebene ist es wichtig, der Datensatz kontinuierlich erweitert wird, um final alle Weine der Welt abzudecken. Dabei ist es jedoch nicht nur bedeutend, dass alle Weine enthalten sind, sondern es sollten auch alle Eigenschaften der Weine bekannt sein. Diese Vollständigkeit der Daten ist beim aktuellen Stand noch sehr dürrig, was sich durch eine Vervollständigung der einzelnen Datensätze schnell verbessern ließ. Darüber hinaus wäre es denkbar, wenn dieser Datensatz für Weinexperten oder Weineinkäufer/Verkäufer mehr als nur eine Indikation geben soll, die Daten weiter zu konkretisieren. So sollten die eins bis fünf Skala, welche es beispielsweise bei der Süße, Säuregehalt und Körper vorhanden ist, durch eine bessere Messskala ersetzt werden. Diese könnte den entsprechenden Zielgruppen mehr Informationen übermitteln als nur ein Wert zwischen eins und fünf.

8. Literatur

1. American Society for Quality (o. J.) What is a Tree Diagram? Systemic or Hierarchy Analysis | ASQ. American Society for Quality. <https://asq.org/quality-resources/tree-diagram>. Zugegriffen: 18. August 2021
2. Beilmann L (o. J.) Alkoholgehalt in Wein - Das solltest du unbedingt Wissen! <https://wein-fuer-laien.de/weinwissen/alkoholgehalt-im-wein/>. Zugegriffen: 19. August 2021
3. Brogsitter Weinversand (o. J.) Gerbstoffe. Brogsitter Weinversand. <https://www.brogsitter.de/weinlexikon/gerbstoffe/>. Zugegriffen: 19. August 2021
4. Brogsitter Weinversand (o. J.) Körper. Brogsitter Weinversand. <https://www.brogsitter.de/weinlexikon/koerper/>. Zugegriffen: 19. August 2021
5. Brogsitter Weinversand (o. J.) Süße. Brogsitter Weinversand. <https://www.brogsitter.de/weinlexikon/suesse/#>. Zugegriffen: 19. August 2021
6. Curran Kelleher (2015) World Countries Hierarchy. GitHub. <https://gist.github.com/curran/1dd7ab046a4ed32380b21e81a38447aa/>. Zugegriffen: 22. August 2021
7. dev7halo (2021) Wine Information. Wine Information with nation, varieties, flavor, price, etc. Kaggle. <https://www.kaggle.com/dev7halo/wine-information>. Zugegriffen: 22. August 2021
8. Fels U (o. J.) Usability-Analyse des Programms Animated Scatter Plot. TU Wien. <https://repositum.tuwien.at/bitstream/20.500.12708/4082/2/Fels%20Ulrich%20-%202015%20-%20Usability%20Analyse%20des%20Programms%20Animated%20Scatter%20Plot.pdf>. Zugegriffen: 25. August 2021
9. Few S (2006) Multivariate Analysis Using Parallel Coordinates. Perceptual Edge. http://www.perceptualedge.com/articles/b-eye/parallel_coordinates.pdf. Zugegriffen: 18. August 2021
10. Petrova M (2002) Web-basierte dynamische Visualisierung klinischer Daten. Ludwig-Maximilians-Universität München. <https://www.nm.informatik.uni-muenchen.de/common/pub/Fopras/petr02/PDF-Version/petr02.pdf>. Zugegriffen: 25. August 2021
11. Roca P (2021) State of the Vitivinicultural World in 2020. OIV Press Conference, o. O.
12. Teufel M (2021) Die richtige Trinktemperatur für Wein - warum wichtig? SWISSCAVE AG. <https://swisscave.com/de/swisscave-blog/post/die-richtige-trinktemperatur-fur-wein-warum-wichtig>. Zugegriffen: 19. August 2021
13. Vineyard99 (2020) Wein & Wissen: Weinjahrgang – Geheimnis der Weinalterung. Vineyard99. <https://www.vineyard99.de/weinjahrgang-und-weinalterung/>. Zugegriffen: 19. August 2021
14. Weinkenner GmbH (2011) Die Säure | Weinkenner.de. Weinkenner GmbH. <https://www.weinkenner.de/die-saeure/>. Zugegriffen: 19. August 2021
15. XE.com Inc. (o. J.) 1 EUR to KRW - Euros to South Korean Won Exchange Rate. XE.com Inc. <https://www.xe.com/currencyconverter/convert/?Amount=1&From=EUR&To=KRW>. Zugegriffen: 08. September 2021
16. Yi M (2019) A Complete Guide to Scatter Plots. Chartio. <https://chartio.com/learn/charts/what-is-a-scatter-plot/>. Zugegriffen: 18. August 2021

Anhang

Git Historie

- * a7d39f2 (HEAD -> main, origin/main) (Korrektur des Berichtes, 2021-09-09)
- * bc38440 (Korrektur Readme, 2021-09-09)
- * 90c6009 (Korrigieren der Webseiten, 2021-09-09)
- * d05eb82 (Korrektur bis Kapitel 3.4, 2021-09-09)
- * 52b5ed5 (Korrigieren komplettes Kapitel 1 - 2, 2021-09-09)
- * 0430c3c (Erweiterung der Eigenschaften Webseite, 2021-09-09)
- * 09a7dc8 (Korrektur des Berichtes, 2021-09-09)
- * 15800a6 (Aktualisierung Bilder, 2021-09-09)
- * 08ecc8e (Aktualisierung der Bilder, 2021-09-09)
- * 4cf85d6 (Hinzufügen von Entern, 2021-09-09)
- * b10ec55 (Einbinden des Eigenschaften Dokumentes, 2021-09-09)
- * 45c000c (Hinzufügen eines Eigenschaften Dokumentes, 2021-09-09)
- * d56bfe5 (Verlinkung der Webseite und Darstellungsanpassung, 2021-09-08)
- * 5bb1a98 (Aktualisierung Webseite Scatterplot, 2021-09-08)
- * ad11d1d (Webseite aktualisieren Baumhierarchie, 2021-09-08)
- * 5c77334 (Aktualisieren Webseite für ParalleleKoordinaten, 2021-09-08)
- * 0135e57 (Löschen von Kommentaren im Code, 2021-09-08)
- * 9f961ae (Optische Anpassung des Baums, 2021-09-08)
- * d2770f5 (Löschen von überflüssigen Dateien, 2021-09-08)
- * 04c4a44 (Änderung des Zitationsstiel, 2021-09-08)
- * 91615eb (Korrektur Kapitel 6 und 7, 2021-09-08)
- * 87bd6ac (Korrektur komplettes Kapitel 5, 2021-09-08)
- * e0f72a7 (Korrektur Kapitel 3.4 und 4, 2021-09-08)
- * 47337af (Korrigieren Kapitel 3.2 und 3.3, 2021-09-08)
- * 60cf2c0 (Korrektur Kapitel 3 und 3.1, 2021-09-08)
- * 3b42be0 (Korrektur Kapitel 2.1 und 2.2, 2021-09-08)
- * 6061665 (Löschen von unnötigem Code, 2021-09-07)
- * ac961d6 (Überarbeitung der Readme, 2021-09-07)
- * 2ae4aae (Korrektur Kapitel 2, 2021-09-07)

- * 18022bb (Korrektur von Kapitel 1.2 und 1.3, 2021-09-07)
- * 6eaf2ee (Korrektur des Kapitels 1.1, 2021-09-07)
- * 74ed78a (Layoutanpassung für den Baum, 2021-09-07)
- * 3c9c6ae (Hovern hinzugefügt, 2021-09-07)
- * aa607f6 (Korrekturlesen von Kapitel 1, 2021-09-07)
- * 95886f8 (Änderung an der Verlinkung, 2021-08-29)
- * 7991904 (Hinzufügen der Seiten für Github Pages, 2021-08-29)
- * 7a06098 (Testen für Github Pages, 2021-08-29)
- * 1baf4f1 (Pflege der Readme, 2021-08-29)
- * d63deb0 (Erstellung eines Files für Github Pages, 2021-08-29)
- * cebd633 (Schriftartanpassung, 2021-08-28)
- * 1fb3535 (Darstellungsanpassungen, 2021-08-28)
- * 0d5d7d2 (Anpassungen in Navigation, 2021-08-28)
- * 154afa5 (Anpassung des Codes Formatierung und Navigation, 2021-08-28)
- * d5954c0 (Workarround für nicht funktionierende Webseiten, 2021-08-28)
- * 7f0a696 (Hinzufügen der Startseite und der Navigation, 2021-08-28)
- * e3d8086 (Hinzufügen andere Webdateien, 2021-08-27)
- * 038aaf9 (Erstellung der Webseite für BaumHierarchie, 2021-08-27)
- * d4d0f35 (Erstellung der JavaScript Dateien, 2021-08-27)
- * 3730550 (Verschiebung in Zwischenstand, 2021-08-27)
- * db588ca (Neuordnung und sortierung für vorbereitung Webseite, 2021-08-27)
- * 3565186 (Anpassung der neune Datenquelle, 2021-08-27)
- * 1f333a6 (Hinzufügen der neuen Datenquelle, 2021-08-27)
- * 29abf40 (Ausprobieren eines andere Lösungsansatzes, 2021-08-27)
- * e531919 (Weiteres Testen wie JS in HTML eingebunden wird, 2021-08-27)
- * 44c8dd9 (Erstellung der JavaScript Dateien, 2021-08-27)
- * d09d68c (Verbesserung der Formatierung, 2021-08-27)
- * 702f511 (Umbenennung des Ordner Aufbereite Daten, 2021-08-27)
- * a1c1d85 (Merge branch 'main' of <https://github.com/RicBre/Elm-Projekt-WineInformation> into main, 2021-08-27)

|\

| * bad7638 (Löschen von elm-stuff, 2021-08-27)

| * 74bc9af (Löschen von .vscode, 2021-08-27)

* | 8cd8b66 (Hinzufügen weiterer Informationen, 2021-08-27)

|/

* c6b7512 (Änderungen Git für bessere Übersichtlichkeit, 2021-08-26)

* d7e4e10 (Codeangleichung, 2021-08-26)

* 02b49b4 (Anpassung der Darstellung, 2021-08-26)

* f221f39 (Strukturierung des Codes, 2021-08-26)

* c2d8eb2 (Anpassung der Darstellungsweise, 2021-08-26)

* 96d0cf2 (Strukturierung des Codes, 2021-08-26)

* c8ed8a3 (Neu Strukturierung und Anpassung der Darstellung, 2021-08-26)

* 6177e06 (Änderungen des Links, 2021-08-26)

* ec87a4a (Umbenennung und verschieben von Dateien, 2021-08-26)

* 6de5f4c (Schreiben des Kapitels 7, 2021-08-25)

* 4204548 (Aktualisierung der Notizen, 2021-08-25)

* 8ab649f (Schreiben des Kapitels 6, 2021-08-25)

* 326a337 (Aktualisierung der Notizen, 2021-08-25)

* 6ff502e (Aktualisierung der Notizen mit potenziellen Studien, 2021-08-24)

* 2fe2ccb (Schreiben von Kapitel 5.3, 2021-08-23)

* c986460 (Screenshoot für Anwendung Baumdiagramm, 2021-08-23)

* ca6c28e (Schreiben des Kapitels 5.2, 2021-08-23)

* e7b969d (Screenshoot Parallele Koordinaten für Anwendung, 2021-08-23)

* e2bce2e (Schreiben des Kapitels 5 und 5.1, 2021-08-23)

* d4f6041 (Screenshoot für Scatterplott mit Preis/Körper, 2021-08-23)

* 576d60f (Aktualisierung der Notizen, 2021-08-23)

* d6bbe2a (Schreiben des Kapitels 4, 2021-08-23)

* 12e8848 (Schreiben des Kapitels 3.4, 2021-08-23)

* 41ef278 (Schreiben des Kapitels 3.3.3, 2021-08-22)

* 90f146d (Aktualisierung der Notizen, 2021-08-22)

* 71e49c2 (Hinzufügen eines Screenshots vom Baumdiagramm, 2021-08-22)

* 84377dd (Schreiben des Kapitels 3.3.1 und 3.3.2, 2021-08-22)

* 480c1a6 (hinzufügen eines Screenshots für den Bericht, 2021-08-22)

- * f59a8b5 (Daten nachvollziehen, 2021-08-20)
- * 88580d6 (Schreiben am Kapitel Präsentation der Visualisierungen - 1, 2021-08-20)
- * 53b15fa (Hinzufügen eines Screenshots für den Bericht, 2021-08-20)
- * a6f29d4 (Fertigstellung des Kapitels Anforderungen an die Visualisierung, 2021-08-20)
- * 6b9af26 (Schreiben des Kapitels 2.2; 3; 3.1, 2021-08-20)
- * 7c23751 (Ergänzung von Notizen in Analyse der Anwendungsaufgaben, 2021-08-20)
- * adaf360 (Suche und Interaktion mit den Daten und Anwendungen, 2021-08-20)
- * 0f17158 (Hinzufügen von Notizen zu Datenvorverarbeitung, 2021-08-20)
- * 5b506d0 (Fertigstellung des Kapitels Daten, 2021-08-19)
- * c9063ad (Fertigstellung des Kapitels Überblick und Beiträge, 2021-08-19)
- * a8c30ef (Fertigstellung des Kapitels Zielgruppen, 2021-08-19)
- * fa8b3aa (Fertigstellung des Teils Anwendungshintergrund, 2021-08-19)
- * a7cf6f7 (Schreiben eines Teils des Anwendungshintergrund, 2021-08-18)
- * a26b123 (Trennung von Notizen und geschrieben Projektbericht, 2021-08-18)
- * dece647 (Schreiben der Einleitung, 2021-08-18)
- * 111ae6a (Filteranpassung an der Excel, 2021-08-18)
- * 49904cd (Hinzufügen der neuen Datenquelle, 2021-08-12)
- * 1bdc20f (Anpassung der Daten der JSON um auf Stand der CSV zu sein, 2021-08-12)
- * 86f39a1 (Fehlerbehebung in den Daten, 2021-08-12)
- * fbc10b0 (Löschen von Datensätzen, 2021-08-12)
- * ec214ed (Hinzufügen der neuen kleineren Daten, 2021-08-12)
- * 54dd79a (Verkleinerung der Daten, 2021-08-12)
- * 21ae0d9 (Anpassung des Codes an neue Ablageorte, 2021-08-12)
- * 428872e (Neuorganisation des Berichtes, 2021-08-12)
- * 3377be5 (Neuorganisation des Codes, 2021-08-12)
- * 2201157 (Strukturierung der Daten in neuen Ordner, 2021-08-12)
- * 8e40230 (Erstellung von Interaktiven Parallelen Koordinaten, 2021-08-11)
- * 74e2f10 (Erstellung eines Interaktiven Scatterplots, 2021-08-11)
- * d5cf947 (Entfernen von Ländern die keine Weine haben, 2021-08-11)
- * 6a5fffb (Hinzufügen des PDFs, 2021-08-09)
- * 93627b2 (Unter Überschriften Inhalte Stichpunktartig ergänzt, 2021-08-08)

- * 288eea4 (Baumhierarchie aktualisierung auf kleineres JSON, 2021-08-08)
- * c7d3e18 (Entfernen aller Länder die keine Weine haben, 2021-08-08)
- * 8546709 (Aktualisierung der Namen in der JSON, 2021-08-08)
- * 36d8b34 (Aktualisierung der Namen für CSV, 2021-08-08)
- * b78036e (Aufbereitung der Namen der Weine, 2021-08-08)
- * b3c52ae (Hinzufügen Wine zu den Ländern, 2021-08-07)
- * a009c49 (Testen von der änderung der Daten, 2021-08-07)
- * 906d590 (Einfügen der neuen Daten für die Baumhierarchie, 2021-08-07)
- * ac10d30 (Datenaufbereitung der Geo Daten, 2021-08-07)
- * cedbc18 (Ausprobieren dynamische Spaltenzuweisung, 2021-08-06)
- * c93707c (Erstellung CSV Decoder für gesamte CSV-Datei, 2021-08-06)
- * 31a1590 (Testlauf für Baumhierarchie, 2021-08-06)
- * 3e6c4fc (Ausprobieren einer neuen Datenquelle und Format, 2021-08-06)
- * bb12ef2 (Versuch CSV in Parallele Koordinaten, 2021-08-04)
- * 48283a2 (Vorbereitungen auf Parallele Koordinaten, 2021-08-04)
- * 0fac7e0 (Datengrundlage wurde aufbereitet, 2021-08-04)
- * e85cb99 (Versuch Scatterplot interaktiv zu gestalten, 2021-08-04)
- * f44e476 (Idee zum Variablen einfügen (nicht vervollständig), 2021-08-03)
- * b9da6ec (Probe zum Interaktiv machen des Scatterplots, 2021-08-03)
- * 9757801 (Erster funktionierender Scatterplot, 2021-08-03)
- * 63f7c19 (Hinzufügen des Punktes, 2021-08-03)
- * 9a774f9 (Anpassung des Scatterplots, 2021-08-03)
- * 3817ccb (Modell für weiteren Float umschreiben, 2021-08-03)
- * 84d2fc6 (Hinzufügen der View zum Scatterplot aus CsvElmTest, 2021-08-03)
- * 74307a0 (CSV Decoder für Scatterplot Kopieren, 2021-08-03)
- * 205e7dc (Versuch Elm-Format in VSC zu installieren, 2021-08-03)
- * 8209e52 (hinzufügen der CSV als Excel, 2021-08-03)
- * 96f59c9 (Versuch gesamte CSV zu Decodieren, 2021-08-03)
- * 854e103 (Hinzufügen weitere Spalte die decoded wird, 2021-08-02)
- * 78290fa (Anbindung der CSV Datei mit GitHub erfolgreich, 2021-07-14)
- * c535a82 (Ausprobieren des einbinden der CSV mithilfe eines GitHub Links, 2021-07-14)

- * 566221c (Hochladen der CSV Dateien um Online Zugriff auf diese zu erhalten (über GitHub), 2021-07-12)
- * 6f249b7 (Ausprobieren des Einbindens der CSV Datei - bisher noch nicht erfolgreich, 2021-07-07)
- * 83029b6 (Word Dokument mit hinzugefügt, 2021-07-07)
- * 756ff9d (Elm initialisiert und Elm.json Datei wurde erstellt, 2021-07-05)
- * 6162ed9 (CSV Dateien zum .gitignore hinzugefügt, 2021-07-05)
- * c08733f (Hinzufügen von .gitignore, 2021-07-04)
- * 533c289 (Initial commit, 2021-07-04)