

Projektbericht zum Modul Information Retrieval und Visualisierung Sommersemester 2022

Thema:

Untersuchung des Einflusses von Alkohol auf Schüler mithilfe
verschiedener Visualisierungstechniken

Vorgelegt von:

Arsela Leskaj

Matrikelnummer: 216104691

Abgabetermin: 21.12.2022

GitHub Repository: https://github.com/Ella199/Elm_Project-Student-Alcohol-Consumption

Projektwebseite: https://ella199.github.io/Elm_Project-Student-Alcohol-Consumption/

Gliederung

1. Einleitung	3
1.1 Anwendungshintergrund	4
1.2 Zielgruppen	6
1.3 Überblick und Beiträge	7
2. Daten	8
2.1 Technische Bereitstellung der Daten	9
2.2 Datenvorverarbeitung	9
3. Visualisierungen	10
3.1 Analyse der Anwendungsaufgaben	10
3.2 Anforderungen an die Visualisierungen	13
3.3 Präsentation der Visualisierungen	15
3.3.1 Visualisierung Eins	15
3.3.2 Visualisierung Zwei	16
3.3.3 Visualisierung Drei	17
3.4 Interaktion	18
4. Implementierung	19
5. Anwendungsfälle	22
5.1 Anwendung Visualisierung Eins	22
5.2 Anwendung Visualisierung Zwei	23
5.3 Anwendung Visualisierung Drei	25
6. Verwandte Arbeiten	26
7. Zusammenfassung und Ausblick	28

1. Einleitung

Eine Studie aus dem Jahr 2007 vom Eurostat statistischem Amt der Europäischen Union (EU), beobachtet frühzeitige Schul- und Ausbildungsabgänger innerhalb der EU. Dabei wird festgestellt, dass die durchschnittliche Schul- und Ausbildungsabgangrate der prozentualen 18- bis 24-Jährigen in der EU im Jahr 2006 bei 15% lag. Portugal stach mit einer sehr hohen Abgangrate von 40% auf.¹

Es ergibt sich die Frage, womit der vorzeitige Schulabbruch zu tun hat. Eine Möglichkeit ist ein zu hoher Alkoholkonsum. Junge Menschen kommen sehr schnell mit Alkohol in Kontakt. Das hat verschiedene Gründe. Soziale Aspekte wie Anerkennung bei anderen Schülern, sich bestimmten Freizeit- oder Lerngruppen zugehörig und verbunden zu fühlen, hängen mit dem Trinkverhalten vieler Jugendlicher zusammen. Viele Schüler lernen in der Gruppe und wollen in wechselseitiger Verbindung mit den anderen Mitschülern stehen, da sie sich gegenseitig austauschen möchten und der Wissenstransfer dabei helfen kann, Fortschritte beim Lernen zu machen und den Unterrichtsstoff besser zu verinnerlichen. Oft trinken manche Jugendliche nach einer Gruppenarbeit, um den erzielten Erfolg von einem langen Lerntag zu feiern. Um sich eine Belohnung zu gönnen, greifen sie zu alkoholischen Getränken. Die familiären Hintergründe spielen auch eine sehr wichtige Rolle, zum Beispiel die oftmals zu hohen Erwartungen der Eltern, die zu Leistungsdruck führen. Diejenigen, die nicht mit dem Lernstress umgehen können sehen das Trinken als einen möglichen Weg, den Stress zu bewältigen.²

Eine Studie der Universität Essex belegt, dass manche Studierende, die nach einem langen Lerntag Alkohol trinken, Informationen besser im Langzeitgedächtnis speichern können als andere, die auf den Alkoholkonsum verzichten.³

Allerdings führt eine regelmäßige übersteigende Trinkmenge an Alkohol zu chronischem Alkoholkonsum, welcher viele Verhaltensstörungen induzieren kann. Solche Verhaltensstörungen können die sozialen Interaktion in der Gruppen beeinträchtigen. Somit kann es zur Reduzierung sozialer Kontakte kommen und das kann wiederum dazu führen, dass manche Schüler von anderen Gruppenmitgliedern ausgeschlossen werden. Viele Studien berichten aber auch, dass der Alkoholkonsum geringere Wirkung auf die Schulnoten hat. Der Alkoholkonsum kann als intrinsische Motivation dienen und sorgt für mehr Interaktion oder weckt das Interesse der

¹ Vgl. Cortez & Silva, Eurostat (2007)

² Vgl. Koler, (2014), Stumpp, Stauber, & Reinl (2009)

³ Vgl. FOCUS online, (2017)

Gruppenmitglieder, mit anderen Mitgliedern Zeit zu verbringen. Darüber hinaus können die Schüler in Lerngruppen zusammenfinden und nach dem Lernen zusammen Zeit verbringen. Alkohol kann auch die soziale Interaktion in der Gruppe stärken. Das Konsumieren des Alkohols kann bestimmte Bereiche im Belohnungssystems des Gehirns simulieren und dementsprechend das Verhalten der Individuen beeinflussen.⁴

Das konkrete Zielproblem, dass mittels Visualisierung gelöst werden soll, ist die Beantwortung der verschiedenen Fragen von unterschiedlichen Zielgruppen in Bezug auf: inwiefern das Alkoholkonsum die Schulnoten beeinflusst und ob andere Merkmale identifiziert werden können die einen direkten Einfluss auf die Endnoten haben. Genauer erklärt wird mithilfe der Visualisierungen versucht zu erkennen, ob eine Beeinträchtigung der Schülerleistungen in Verbindung mit Alkoholkonsum vorliegt und ob das Alkoholkonsum in Kombination mit anderem u.a. sozialen Merkmalen usw. die Noten in der Schule beeinflusst.

Im Fokus dieser Projektarbeit steht auch die Untersuchung der Zusammenhänge zwischen dem Alkoholkonsum und den erreichten Schulnoten von zehnter bis zwölfter Klasse anhand von drei Visualisierungstechniken, die im Kapitel 1.1 vorgestellt werden. Darüber hinaus werden in dieser Projektarbeit interessante Erkenntnisse zu der Thematik geliefert. Weiterhin werden auch andere demographische (zum Beispiel Bildungsniveau der Eltern) und soziale (beispielsweise Freizeit nach der Schule) Merkmale in Erwägung gezogen um zu verstehen, ob noch andere Merkmale die Schülerleistungen beeinflussen.

1.1 Anwendungshintergrund

Ziel dieser Arbeit ist es, eine Übersicht anhand von drei ausgewählten Visualisierungstechniken zum Thema “Untersuchung des Einflusses von Alkohol auf Schüler mithilfe verschiedener Visualisierungstechniken” zu bieten. Die verwendeten Techniken sind Scatterplot, Parallele Koordinaten und Stickfigureplot. Es ist anzumerken, dass in diesem Kapitel nur eine Überblick zu den konkreten verwendeten Visualisierungstechniken gegeben soll. Die konkreten Techniken werden aus der Analyse und den Anforderungen in Kapitel 3 abgeleitet und ausführlich erklärt.

Ein Scatterplot bildet die Datenpunkte in einem zweidimensionalen Raum ab. Mithilfe dieser Visualisierungstechnik lässt sich die Beziehung zwischen Attributwerten auf den Koordinatenachsen untersuchen.

⁴ Vgl. Soyka (2001), Doerfel (2018)

So können beispielsweise Informationen über zwei numerische Beträge - zum Beispiel Alkoholkonsum unter der Woche und Abschlussnoten aus dem Fachbereich Mathematik - geliefert werden. Diese Technik beinhaltet weitere Vorteile, wie zum Beispiel die Identifizierung bestimmter Muster in den Datensätzen. Somit lassen sich Zusammenhänge zwischen Datensätzen untersuchen und es können Aussagen gemacht werden, ob zwischen den Variablen unter anderem eine positive oder negative Korrelation besteht. Weiterhin können die Anwender zu neuen Erkenntnissen gelangen und interessante Rückschlüsse zu diesem Thema ziehen, indem sie anhand dieser Visualisierung die Möglichkeit erhalten, Ausreißpunkte in den Datensätzen zu identifizieren.⁵

Die zweite Visualisierung innerhalb dieser Forschungsarbeit sind die Parallele Koordinaten. Bei dieser Technik besteht die Möglichkeit, die Datensätze in einem mehrdimensionalen Raum darzustellen. Die Koordinatenachsen dieser Visualisierung bauen sich parallel zueinander auf und stellen die Werte für vier verschiedene numerische Variablen dar. Attributwerte werden durch Linien dargestellt. Vier numerische Beiträge der jeweiligen Variablen lassen sich gleichzeitig anzeigen und durch Anklicken bestimmter Variablen kann man die Linien eindeutig identifizieren. Somit lassen sich einzelne Zusammenhänge zwischen den Attributen feststellen. Weiterhin lässt sich beobachten, wie eng sich Liniengruppen zusammenschließen und in welche Richtung sie sich bewegen.⁶

Die dritte und somit letzte Technik, die zur Visualisierung der Datensätze verwendet wurde, ist Stickfigureplot. Durch diese Technik lassen sich die Daten in einem mehrdimensionalen Raum darstellen. Diese Technik zieht im Vergleich zu den anderen vorgestellten Techniken mehr Variablen in Betracht. Eine von den Eigenschaften dieser Technik sind unter anderem wahrnehmungsbasierte Darstellungen und gleichzeitige Beobachtung der Zusammenhänge und Muster zwischen sieben Variablen. Ähnlich wie bei Scatterplot können hier auch Korrelationsbeziehungen und Ausreißer in den Datensätzen identifiziert werden.⁷

Die für diese Projektarbeit verwendeten Datensätzen stammen aus zwei öffentlichen Sekundarschulen: Gabriel Pereira und Mousinho da Silveira in Portugal. Die Datensätze wurden in den Jahren 2005 und 2006 erhoben. Das Schulsystem in Portugal besteht aus neun Jahren Grundschulbildung, gefolgt von drei Jahren Sekundarschule. Wobei die neun Jahre Grundschulbildung die Voraussetzung für die Sekundarschule sind. Die Bewertungsskala in der

⁵ Vgl. Rumsey (2010)

⁶ Vgl. Gemignani (2021)

⁷ Vgl. Pickett & Grinstein (1988)

Sekundarschule reicht von 0 bis 20 Punkten. 0-9 Punkte gelten als nicht bestanden und 20 ist die beste Note. In den Datensätzen werden die Schulnoten von der zehnten bis zur zwölften Klasse für die Schulfächer Mathematik und Portugiesisch erfasst. Die Abschlussnoten der zwölften Klasse ergeben sich aus dem Durchschnitt der zehnten und elften Klasse. Die Datensätze enthalten wesentliche Informationen unter anderem über schulbezogene Leistungen sowie soziale und demographische Merkmale. Somit können anhand der Analyse bestimmter Variablen in den Datensätzen Erkenntnisse hinsichtlich der Beeinträchtigung der Schülerleistungen in den beiden Schulfächern gewonnen werden.⁸

1.2 Zielgruppen

Für die vorliegenden Daten sind im Rahmen dieser Arbeit drei potenzielle Zielgruppen denkbar: Data Science-Unternehmen, Schuldirektoren und Sozialforschung.

Hierbei ist anzumerken, dass Data Science-Unternehmen sich mit den Themen innerhalb öffentlichen Sektor insbesondere Bildungsforschung in der Schulen beschäftigen. Sie sind als Zielgruppe möglich, um konkrete Vorhersagen zu treffen hinsichtlich der Abschlussnote und um herauszufinden, ob die Absolventen nach einem universitären Hochschulabschluss oder einer betrieblichen Ausbildung streben. Zudem können sie auch langfristige Untersuchungen zur Bevölkerungsentwicklung oder einen Generationsvergleich in Portugal machen.

Auch für die Schuldirektoren können die Ergebnisse dieser Forschungsarbeit von Interesse sein. Beispielsweise können sie sich dafür interessieren, wann welche Präventionsmaßnahmen eingeführt werden sollten, um die Leistung der Schüler zu verbessern. Ein Schuldirektor kann sich auch für eine bessere Lernqualität interessieren und die entsprechenden Maßnahmen treffen, um dies zu ermöglichen.

Anhand der Daten kann erkannt werden, ob die Schüler in naturwissenschaftlichen (Mathematik) oder in geisteswissenschaftlichen (Portugiesisch) Fächern am besten abschneiden. Zudem ist es für den Schuldirektor interessant, wann die Schüler am meisten Alkohol trinken - am Wochenende oder unter der Woche - und wie der Alkoholkonsum mit den Schulnoten zusammenhängt. Denkbar wäre, Unterrichtsstunden für das Thema Alkohol einzuführen, damit sich die Schüler besser informieren können. Wobei zu beachten ist, dass das nur eine Empfehlung ist.

In Kapitel 3 wird anhand der Visualisierungen in Erfahrung gebracht, ob die Schüler ein Problem mit Alkohol haben. Alternativ wäre auch möglich für die Schüler, die mehr als durchschnittlich Alkohol konsumieren, Coaching in der Schule einzuführen.

⁸ Vgl. Cortez & Silva (2008), Taborda (2022)

Eine weitere potenzielle Zielgruppe ist die Sozialforschung. Sie kann untersuchen, ob sozialstrukturelle Aspekte den Alkoholkonsum beeinflussen, zum Beispiel inwieweit der Sozialstatus der Eltern die Schulnoten beeinflusst. So kann die Sozialforschung in Erfahrung bringen, ob diejenigen Kinder, die aus einer Familie stammen, in der die Eltern ein akademisches Bildungsniveau verfügen, bessere Noten haben.

1.3 Überblick und Beiträge

Das Thema Alkoholkonsum bei Jugendlichen wird weltweit diskutiert und es werden Studien beispielsweise an Schulen durchgeführt, welche den Zusammenhang zwischen Schulleistung und Alkoholkonsum untersuchen.⁹

Mithilfe dieser Projektarbeit können mögliche Fragen, die das Interesse der angesprochenen Zielgruppen zu diesem Thema entsprechen, beantwortet werden. Die Projektarbeit beschäftigt sich mit folgenden Fragen: Werden die Abschlussnoten von den früheren erreichten Noten aus der zehnten und elften Klasse beeinflusst? Inwieweit beeinträchtigt der Alkoholkonsum die erreichten Schulnoten der zehnten, elften und zwölften Klasse in den Fächern Mathematik und Portugiesisch? Welche Noten erreichen die Schüler bezogen auf das Ausmaß ihres Alkoholkonsums? Wie kann eine Kombination des Alkoholkonsums mit demographischen und soziablen Variablen die Schülerleistung beeinflussen? Wie sehen die geschlechtsspezifischen Unterschiede hinsichtlich des Zusammenhangs von Alkoholkonsum und der Abschlussnoten aus?

Um viele Interessenten darüber zu informieren und einen guten Überblick zu bieten, wurden in dieser Projektarbeit drei Visualisierungstechniken verwendet: Scatterplot, Parallele Koordinaten und Stickfigureplot. Scatterplot ermöglicht den Nutzern eine zweidimensionale Darstellung der Datensätze. Um diese Technik anzuwenden, wird von den Anwendern kein Vorwissen im Fachbereich verlangt. Scatterplot stellt eine benutzerfreundliche Anwendung für die Nutzer dar.

Parallele Koordinaten ist eine weitere Visualisierungstechnik, die in dieser Arbeit angewendet wird. Bei dieser Technik können beliebige Variablen ausgewählt und gegenüber gestellt werden. Somit können wesentliche Informationen aus den Datensätzen gezogen werden. Diese Art von Visualisierung fordert keine Vorkenntnisse der Anwender und ist ebenfalls benutzerfreundlich. Im Gegensatz zu Scatterplot weist diese Technik mehr Interaktionsmöglichkeiten zwischen den Variablen auf, da vier Attributwerte ausgewählt werden können.

Bei der dritten Visualisierungstechnik handelt es sich um einen Stickfigureplot und diese Technik weist eine mehrdimensionale Datendarstellung auf. Besonders bei dieser Darstellung ist, dass

⁹ Vgl. Balsa, Giuliano & French (2011), DeSimone & Wolaver (2005), Dee & Evans (2003)

gleichzeitig sieben Variablen beobachtet werden können. Im Vergleich zu den anderen Visualisierungstechniken wurden bei dieser Technik Variablen - unter anderem soziale Attributwerte - dargestellt, die nicht in den vorherigen zwei Visualisierungstechniken vorkommen. Um diese Technik anzuwenden, ist es von Vorteil, über Vorkenntnisse zu der Thematik zu verfügen.

2. Daten

Für die bereits erwähnten Visualisierungstechniken werden Daten aus der Online-Plattform Kaggle entnommen. Die Daten dienen ursprünglich einem wissenschaftlichen Artikel, der sich mit der Methoden des Data Mining beschäftigt und Vorhersagen über die Schulnoten der Sekundarschüler in Portugal macht.¹⁰ Ähnliche Arbeiten können bei der Suchmaschine Google gefunden werden. Einige Studien werden im Kapitel 6 als Vergleich zu dieser Projektarbeit genutzt um tiefergreifende Einblicke über die verwendeten Visualisierungstechniken zu liefern.

Die zur Verfügung gestellten Datensätze bei der Plattform Kaggle bestehen aus zwei CSV-Dateien und einer weiteren Datei, die dazu dient, die zwei CSV-Dateien in einer Datei zusammenzuführen.

Die eine CSV-Datei beinhaltet die erhobenen Daten für das Schulfach Portugiesisch und besteht insgesamt aus 649 Datensätzen. Die andere CSV-Datei für das Schulfach Mathematik enthält insgesamt 395 Datensätze. Aus diesem Grund wurde entschieden, den zur Verfügung gestellten R-Code (student-merge.R) bei der Webseite Kaggle auszuführen, um einen einheitlichen Datensatz der beiden Studienfächer zu bekommen. Nach dem Ausführen der mitgelieferten R-Anweisungen von den Autoren, wurde die ursprüngliche CSV-Datei mit 382 Datensätzen verkleinert. Die neue CSV-Datei beinhaltet die gemeinsamen Angabedaten derjenigen Studenten, die die beiden Umfragen sowohl für Mathematik als auch für Portugiesisch ausgefüllt haben. Eine wichtige Eigenschaft der Datensätze besteht darin, dass sich die Schüler nicht eindeutig identifizieren lassen, da die Daten anonym erhoben sind. Aus diesem Grund wurde für die Zusammenführung der beiden Datensätze in die neue CSV-Datei nach identischen Merkmalen der Schüler gesucht, die jeden Schüler charakterisieren.¹¹ Die Attributen, die für die Visualisierungstechniken verwendet wurden, sind die Schulnoten von zehnte, elfte und zwölfte Klasse, das Ausmaß des Alkoholkonsums sowohl am Wochenende als auch unter der Woche, Fehltage in der Schule, Freizeit nach der Schule, das Bildungsniveau der Eltern und das Geschlecht von Schülerinnen und Schülern. Das Ausmaß des Alkoholkonsums schwankt von Stufe eins bis fünf, wobei Stufe eins für den geringsten Alkoholstufe steht und fünf für den höchsten Alkoholstufe steht. Das Bildungsniveau der Eltern

¹⁰ Vgl. Cortez & Silva (2008)

¹¹ Vgl. Student Alcohol Consumption, update (2016)

umfasst die Stufe null steht für keine Bildung, eins für Grundschulbildung, zwei für 5. bis 9. Klasse, drei für Sekundarschulbildung und vier für Hochschulbildung.

Die Datensätze lassen sich mit den ausgewählten Visualisierungstechniken gut anwenden und die Zielgruppen können daraus wesentliche Informationen in Erfahrung bringen. Auf das Thema wird im Kapitel 5 ausführlich eingegangen und Praxisbeispiele in Bezug auf die Zielgruppen geben.

2.1 Technische Bereitstellung der Daten

Für die Aufbereitung der Daten wurde die Programmiersprache Python verwendet. Python wurde benutzt, um mögliche Fehler bei den Datensätzen zu vermeiden, damit verlässliche Ergebnisse erzeugt werden können. Zudem wurden mit Hilfe von Python einige Datensätze ausgelassen, die für die Visualisierungstechniken nicht geeignet waren. Insgesamt gab es, wie in dem Kapitel zuvor erwähnt, 649 Befragte, die die Daten vollständig hinsichtlich der Sprache Portugiesisch und 382 für Mathematik eingegeben haben. Nach der Datenzusammenführung wurden die Daten mit Hilfe der Programmiersprache Python bearbeitet. Die Daten befinden sich in einem Ordner Data, der aus zwei unterschiedlichen Unterordnern besteht und zwar aus CSV und Pandas. Die Originaldaten sind im Ordner CSV zur Verfügung gestellt, damit die Anwender die Datensätze jederzeit anschauen können, um ein besseres Verständnis der Thematik zu erlangen. Zudem ist es auch wichtig, dass die Nutzer der Github “Student Alcohol Consumption” den Zugang zu den Daten dieser Forschungsarbeit erhalten, um sowohl die Datenstruktur der originalen Datensätze als auch die aufbereiteten Datensätze zu haben.¹²

2.2 Datenvorverarbeitung

Für die Verwaltung und Analyse von Daten wurde die Software-Bibliothek Pandas von Python verwendet. Mithilfe dieser Programmibibliothek wurden die Datensätze in mehreren Schritten verarbeitet. Bevor die Daten auf Github hochgeladen wurden, wurden einige Schritte vorgenommen, die notwendig für eine Datenbereinigung sind.

Diese Schritte umfassen die Umbenennung, das Entfernen von doppelten Einträgen und NaN-Werten (Not a Number/keine Zahl). Weiterhin wurde mithilfe dieser Programmiersprache eine Analyse hinsichtlich des Datentyps durchgeführt.

Nach dem Zusammenführen der beiden Datensätze in eine gemeinsame CSV-Datei wurden bei den identischen Attributen jeweiligen Schülern automatisch neue Spalten .x für das Schulfach

¹² Vgl. Student Alcohol Consumption, update (2016)

Mathematik und .y für das Schulfach Portugiesisch in den Datensätzen hinzugefügt. In einem weiteren Schritt wurden doppelte Einträge entfernt und nur die identischen Attribute beibehalten. Weiterhin wurden die Schulnoten für die zehnte, elfte und zwölfte Klasse mit der jeweiligen Schulfachbezeichnung umbenannt. Die Analyse hinsichtlich der Datentypen der Attribute ist für die nächsten Schritte bezüglich der Programmierung der Visualisierungstechniken wichtig. Darüberhinaus können Informationen geliefert werden, welche Attribute unter anderem einen numerischen- oder einen Text-Wert zuweisen.

Für das Einlesen der Daten in die Elm-Programmiersprache wurde entschieden, den Dateityp CSV zu verwenden, weil sich die Werte des Dateityps Excel in dieser Programmiersprache nicht einlesen konnten. Der Datensatz, der für die Bearbeitung der vorgestellten Visualisierungstechniken verwendet wurde, wurde als mergedstudent_FINAL_NaN.CSV im Ordner Daten/CSV gespeichert.

3. Visualisierungen

Das folgende Kapitel bietet einen umfassenden Überblick über die drei Visualisierungstechniken, die Anforderungen der Zielgruppen und die jeweiligen Interaktionsmöglichkeiten. Dieses Kapitel besteht aus vier Haupt-Unterkapiteln: Analyse der Anwendungsaufgaben, Anforderungen an die Visualisierungen, Präsentation der Visualisierungen und Interaktion. Wobei das Haupt-Unterkapitel namens Präsentation der Visualisierungen aus drei weiteren Unterkapiteln - Visualisierung Eins, Zwei und Drei - besteht.

3.1 Analyse der Anwendungsaufgaben

Die bereits erwähnten Visualisierungen sollen den Anwendern dabei helfen, interessante Erkenntnisse zu gewinnen und aussagekräftige Rückschlüsse zu ziehen. Hier werden die Anwendungsaufgaben analysiert, die für die Lösung des Zielproblems bzw. der verschiedenen Forschungsfragen aus unterschiedlichen Zielgruppen durch die Anwendern bearbeitet werden müssen. Die erste Visualisierung Scatterplot beantwortet folgende Forschungsfragen: Welche Noten erreichen die Schüler bezogen auf das Ausmaß ihres Alkoholkonsums? Inwieweit beeinträchtigt der Alkoholkonsum die erreichten Schulnoten der zehnten, elften und zwölften Klasse in den Fächern Mathematik und Portugiesisch? Die Visualisierung der Datensätze in einem Scatterplot kann von den Anwendern dazu verwendet werden um zwei numerische Attribute miteinander zu vergleichen. Neben den numerischen Attributen wird auch ein Text Attribut namens "sex" angezeigt, um die Schüler geschlechtsspezifisch zu unterscheiden. Die Anwender können auf den ersten Blick

erkennen, um welches Geschlecht es sich in der Visualisierung handelt, da die Schülerinnen und Schüler durch zwei Farben, rot für Schülerinnen und blau für Schüler, deutlich unterscheiden lassen. Da die Angabe der Daten anonym durchgeführt wurde, lässt sich nicht erkennen, welche Schüler welche Daten angegeben haben. Bei der Scatterplotvisualisierung können beliebige Attribute miteinander kombiniert werden. Wie bereits im vorherigen Kapitel erwähnt, sind Vorkenntnisse bei dieser Art der Visualisierung nicht erforderlich.

Es wird aber empfohlen, sich mit dem Thema Alkoholkonsum von Schülern auseinander zu setzen, bevor die Visualisierungstechniken angewendet werden können, damit die Kombination der Attribute sinnvoll ist. Diese Herangehensweise ist wichtig, um die entsprechende Frage hinsichtlich der Untersuchung zu definieren. Somit können Erkenntnisse hinsichtlich der Zusammenhänge zwischen zwei beliebig ausgewählten Attributen abgeleitet werden. Mithilfe der Scatterplotvisualisierung können neue Erkenntnisse zu spezifischen Untersuchungszwecken gewonnen werden, indem die Anwender selbst die Auswahl beziehungsweise die Kombination der Variablen festlegen.

Nutzern soll es möglich sein, einen fundierten Einblick in das Thema zu bekommen, indem sie Zusammenhänge zwischen den Variablen vergleichen können. Zum Beispiel können sie die zwei Beobachtungen - beispielsweise den Zusammenhang der Schulnoten der zwölften Klasse in den Schulfächern Mathematik und Portugiesisch - interaktiv vergleichen.

Diese Analyse ist empfehlenswert, weil die Anwender dadurch nicht nur allgemeine Zusammenhänge zwischen Alkoholkonsum und Noten untersuchen können, sondern auch die Besonderheiten des Einflusses des Alkoholkonsums auf die Schulleistung in den genannten Schulfächern. Darüberhinaus lassen sich Muster aufdecken und Trends identifizieren.

Durch die beliebige Auswahl von Kombinationen der Attribute können viele Fragen der Zielgruppen beantwortet werden. Es können beispielsweise Untersuchungen durchgeführt werden, wie der Alkoholkonsum die Noten der Schulfächer Mathematik oder Portugiesisch beeinträchtigt oder ob der Alkoholkonsum zur Verbesserung der Schulnoten führt. Die Auswahl der Schulfächer kann auch die Frage beantworten, ob der Alkoholkonsum einen schlechten beziehungsweise keinen oder sogar einen positiven Einfluss auf geisteswissenschaftliche oder naturwissenschaftliche Fächer hat. Interessieren sich die Zielgruppen auch dafür, wie regelmäßig Alkohol eingenommen wird, ob unter der Woche oder am Wochenende, können die entsprechenden Attribute aus der x- und y-Achse ausgewählt werden, um mehr über das Thema zu erfahren.

Mit Hilfe der Visualisierungstechnik Scatterplot lassen sich zwei Attributwerte in einem zweidimensionalen Raum vergleichen.

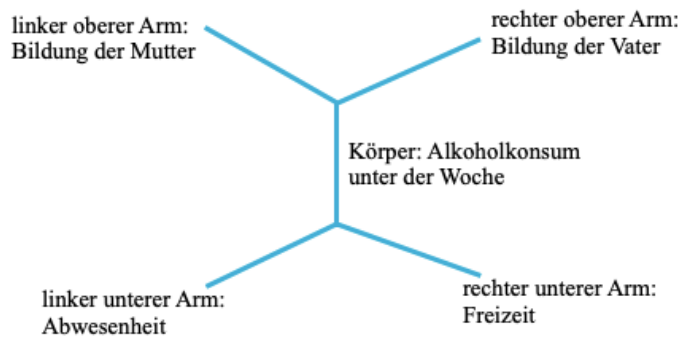
Die zweite Visualisierung Parallele Koordinaten beantwortet folgende Forschungsfragen: Werden die Abschlussnoten von den früheren erreichten Noten aus der zehnten und elften Klasse beeinflusst? Wie sehen die geschlechtsspezifischen Unterschiede hinsichtlich des Zusammenhangs von Alkoholkonsum und der Abschlussnoten aus?

In den Parallele Koordinaten können die Anwender mehr als zwei Attribute miteinander in Bezug setzen. Die Anwender können mithilfe dieser Visualisierungstechnik vier Attribute auswählen und Zusammenhänge zwischen den ausgewählten Variablen untersuchen. Hier können die Anwender die geschlechtsspezifischen Unterschiede durch die vordefinierten Farben intuitiv erkennen. Darüberhinaus steht die Farbe Blau für männliche Schüler und rot für weibliche Schüler.

Interessieren sich Anwender für die Untersuchung bestimmter Eigenschaften, können auf vier Achsen die Werte ausgesucht und die Ergebnisse durch mit Linien miteinander verbundene Attribute angezeigt werden. Sobald die Attributwerte aus vier Achsen ausgewählt sind, wird die Linie mit der Frage grün markiert und das Geschlecht der Schüler sowie die Namen der jeweiligen Werte aus vier ausgewählten Variablen angezeigt. Durch diese Visualisierungstechnik können mehrere Trends und unterschiedliche Zusammenhänge in den ausgewählten Werten gefunden werden.

Die dritte Visualisierung Stickfigureplot beantwortet folgende Forschungsfragen: Wie kann eine Kombination des Alkoholkonsums mit demographischen und soziablen Variablen die Schülerleistung beeinflussen?

Stickfigureplot ermöglicht eine mehrdimensionale Darstellung der Datensätze. Ein Stickfigure ist aus fünf Bestandteilen aufgebaut, und jeder Bestandteil ist einem numerischen Wert zugeordnet. Im Vergleich zu den anderen vorgestellten Visualisierungstechniken werden hier neben den Schulnoten beider Fächern und dem Alkoholkonsum noch weitere Attribute analysiert. Diese Untersuchung soll den Zielgruppen insbesondere Sozialforschung dabei helfen, die Forschungsfrage zu beantworten neue und relevante Zusammenhänge zwischen den Variablen zu erkennen. Wie bereits erwähnt besteht eine Stickfigur aus fünf Eigenschaften: Körper, rechter oberer Arm, linker oberer Arm, rechter unterer Arm und linker unterer Arm. Dem Körper ist der Attributwert der Bildung der Mutter, dem rechten oberen Arm der Attributwert der Bildung des Vaters, dem linken oberen Arm der Attributwert der Freizeit der Schüler, dem rechten unteren Arm der Attributwert der Abwesenheit von der Schule und dem linken unteren Arm der Attributwert des Alkoholkonsums unter der Woche zugeordnet.



Die Anwender können durch Anklicken einzelner Stickfigur die Attributwerte im Detail betrachten. Wobei zu beachten ist, dass durch beliebiges Auswählen der x- und y-Achse eine unregelmäßige Verteilung der Punkte entstehen kann, und somit die Texturen schwer zu erkennen sind. Um dieses Problem zu beheben, wurde eine Linie aufgebaut, die von links nach rechts gezogen werden kann, um die Länge der Liniensegmente jeder Stickfigur zu bestimmen. Dabei empfiehlt es sich, die Länge sieben auszuwählen, um die einzelnen Plots besser zu erkennen. Außerdem lässt sich diese Aussage nicht verallgemeinern, da die Anwender selbst bestimmen können, mit welchen Längen sie gern arbeiten möchten.

3.2 Anforderungen an die Visualisierungen

In Kapitel 3.1 wurden die Anwendungen der jeweiligen Visualisierungstechniken beschrieben und diesbezüglich sollen die zu erfüllenden Anforderungen in diesem Abschnitt definiert werden. Den Anwendern soll es möglich sein von der Hauptseite zu den verschiedenen Visualisierungen zu gelangen. Darüberhinaus soll eine Navigationsleiste aufgebaut werden, die die Navigation zu den verschiedenen Darstellungen ermöglicht. Nachfolgend werden die Anforderungen für Visualisierungen aufgestellt werden, die sich aus der Analyse in der vorherigen Kapitel ergeben haben. Es ist anzumerken, dass die Anforderungen für jede Visualisierung einzeln beschrieben werden. Da die verschiedenen Forschungsfragen zu den jeweiligen Visualisierungstechniken eingeordnet wurden, fallen die Anforderungen für jede Visualisierung unterschiedlich aus.

Die Anforderungen für die Visualisierung der Daten in einem Scatterplot ergeben sich aus dem Vergleich von zwei verschiedenen Attributwerten. Dabei soll dementsprechend eine Auswahl der Attributwerten in der Scatterplot dynamisch aufgebaut werden. Somit sind die Zielgruppen, in der Lage individuell zu entscheiden, welchen Attributwert sie miteinander vergleichen möchten. Um möglichst viele Zusammenhänge in der Daten zu erkundigen, soll den Anwendern eine große Auswahl an Variablen zu Verfügung gestellt werden. Weiterhin soll für Anwender erkennbar sein,,

ob die Daten übereinander liegen. Darüberhinaus sollen die übereinander liegenden Kreise durch Betonung der Farbe stärker hervorgehoben werden, damit die Anwender verstehen können, dass auch andere Punkte auf der selben Position in den Achsenkoordinaten stehen. Um geschlechterspezifische Merkmale in der Visualisierung zu analysieren, sollen die Kreise in zwei Farben Rot und Blau aufgeteilt werden. Die ausgewählten Kreise sollen durch Anklicken farblich markiert und mit darüber schwebenden Werten angezeigt werden. Im Vergleich zu Scatterplot sind die Anforderungen bei Parallele Koordinaten wesentlich anspruchsvoller, da die Nutzer durch das Wechseln der vier Achsen, verschiedene Zusammenhänge in den Parallele Koordinaten erkunden können. Die Visualisierung soll so aufgebaut werden, dass die Anwender die Auswahl der Achsen selbst bedienen können und aus den Daten Zusammenhänge erkennt. Darüberhinaus sollen Buttons und das Dropdown-Menü übersichtlich beschriftet sein und das Geschlecht der Schüler deutlich durch Farben identifiziert sein. Hinsicht der Geschlecht der Schüler wird die Farbe Blau für Schüler und die Farbe Rot für Schülerinnen verwendet. Somit können Anwender in den Parallele Koordinaten durch Anklicken der Buttons und Dropdown-Menü selbst festlegen, beispielsweise welche Schulnoten sie aus welchen Klassen in Zusammenhang mit dem Alkoholkonsum - entweder unter der Woche oder am Wochenende - vergleichen möchten. Die vier ausgewählten Attribute sollen durch eine farbig hervorgehobene Linien mit darüber liegenden Werten verbunden und dargestellt werde. Diese Anwendung soll benutzerfreundlich gestaltet werden, damit es den Anwendern leicht fällt, die Attribute auszuwählen und somit x-Einheiten auf der x-Achse und y-Einheiten auf der y-Achse zu identifizieren.

Zu den Anforderungen bei den Stickfigureplots gehört unter anderem der Vergleich der Attribute, wobei die Interessenten die Möglichkeit haben im Gegensatz zu anderen Techniken mehrere Merkmale der Daten gleichzeitig analysieren zu können. Neben dieser Visualisierung soll ein Stickfigur als Demonstration visuell dargestellt werden und die einzelnen Körperteile sollen deutlich beschriftet werden. Zudem soll den Anwendern möglich sein die Größe des Stickfigures mit dem Regler so einzustellen, dass eine Textur sichtbar wird. Weiterhin soll eine Dropdown-Menü aufgebaut werden, damit die Klassenstufen, um die Mathematik- und Portugesisch-Noten mit weiteren Daten in der Stickfigur-Darstellung zu erkunden. Darüberhinaus soll für die Anwender möglich sein, durch Auswählen der Attribute in der Dropdown-Liste, die Verschiebung der visualisierten Stickfigur zu beobachten, um Aussagen über die Signifikanz der Werte machen zu können.

Bei der Auswahl mit der Maus sollen Stickfiguren farblich angezeigt und die numerischen Variablen sollen über den Stickfigur übersichtlich dargestellt werden. Da in dieser Projektarbeit nicht davon

ausgegangen werden kann, dass Schuldirektoren und die Sozialforschung die benötigten Vorkenntnisse haben, werden die Datenvisualisierungen übersichtlich und verständlich dargestellt. Somit soll die Beschriftung der Buttons, Dropdown, Navigationsliste, Achsen, Stickfigures, Punkte und Linien überschaubar dargestellt und farblich markiert werden.

Zudem verfügen alle Visualisierungen über Dropdown und eine Navigationsleiste, die deutlich beschriftet sind. Mithilfe der Navigationsleiste können die Anwender zu den beliebigen

Visualisierung durch das Anklicken der Navigationsleiste ansteuern. Es ist dabei anzumerken, dass aus Gründen des Platzmangels die Beschriftungen abgekürzt benutzt werden. Die Bedeutung jeder verwendeten Abkürzungen wird in der Datei readme.md erklärt. Die Abkürzungen der jeweiligen Variablen wurde sorgfältig ausgewählt, sodass die Anwender sich intuitiv zurecht finden. Somit können die Anwender instinktiv die Dropdown-Liste auswählen, um die x-und y-Achsen mit verschiedenen Attributen zu verwenden. Sobald eine Eigenschaft aus der Dropdown-Liste ausgewählt wird, werden die x-und y-Achsen automatisch beschriftet. Die Beschriftung der Achsen ist optisch gut lesbar. Um die Komplexität der Visualisierungen mit Interaktionsmöglichkeit benutzerfreundlicher darzustellen, wurden die Schüler für alle Visualisierungen in männlich und weiblich durch zwei Farben unterschieden.

3.3 Präsentation der Visualisierungen

In diesem Unterkapitel werden alle Visualisierungstechniken dargestellt und anhand konkreter Beispiele genauer beschrieben.

3.3.1 Visualisierung Eins

Wie in der Abbildung 1 zu sehen ist, können die Anwender beliebig und intuitiv den Dropdon-Menü bedienen und sich schnell und übersichtlich die zugrundeliegende Daten aufzeigen lassen.

Die Abbildung zeigt einen ausgewählten Kreis, der farblich markiert ist. Zu dem Kreis gehören die Attributwerte vom Geschlecht des Schülers, Abschlussnote in Mathematik und Alkoholkonsum unter der Woche. Zudem können Trends beziehungsweise Veränderungen der Entwicklungsrichtung der Daten untersucht werden um daraus Handlungsempfehlungen abzuleiten. Ein wesentlicher Vorteil dieser Visualisierungstechnik besteht darin, dass x- und y-Achsen dynamisch und automatisch angepasst werden können. Diese Visualisierung ist optisch ansprechend für Eine zweidimensionale Darstellung der Daten. Eine alternative zu Scatterplot wäre eine Box plot Darstellung. Allerdings ist in dieser Arbeit eine Scatterplot Darstellung in Vergleich zu einer Box

plot Darstellung besser geeignet, da die Anwender die Beziehung zwischen zwei Attributen besser betrachten und mögliche Trends in der Datensätzen erkennen können.

Alkoholkonsum (Wochentag)

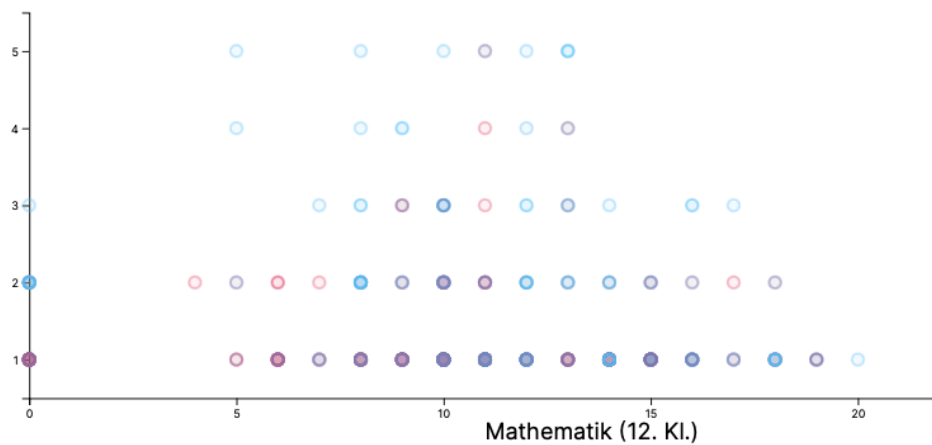


Abbildung 1: Scatterplot (Quelle: eigene Darstellung)

3.3.2 Visualisierung Zwei

Die Anwender können bei der Visualisierung Parallele Koordinaten, neben der Dropdown-Liste Buttons auswählen, um beliebig viele Darstellung der Daten aufzuzeigen. Ein wesentlicher Vorteil dieser Technik ist, dass die Anwender gleichzeitig vier Attribute auswählen können und die Anzahl der Merkmale dynamisch anpassbar machen können. Somit lassen sich mehrere Zusammenhänge von Attributen analysieren und neue Erkenntnisse gewinnen.

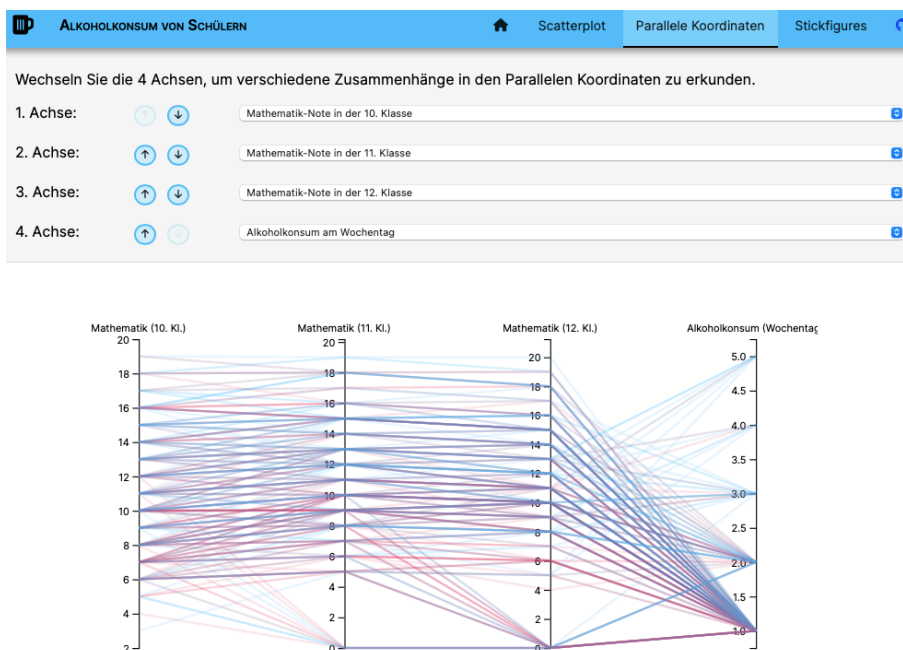


Abbildung 2: Parallele Koordinaten (Quelle: eigene Darstellung)

Wie in der Abbildung 2 zu sehen ist, können beliebige Variablen durch Mausklick ausgewählt werden. Nach dem auswählen der Werte wird langsam in die Visualisierung eine Linie mit rein gezeichnet. Diese Linie ist farblich markiert, damit sie von den anderen Farben eindeutig unterschieden werden kann. Die Anwender können die einzelnen Werte der vier gewählten Attribute schnell ablesen. Auf diesem Weg lassen sich Informationen schneller erkennen und einfacher verstehen. Alternativ zu dieser Darstellung kann eine Liniengraph visualisiert werden. Jedoch ist diese Visualisierungsart nicht gut geeignet, weil durch diese Darstellung die erhobenen Datensätze unübersichtlich sein würden. Die Anwender würde es schwer fallen Rückschlüsse aus der Daten zu ziehen. Darüberhinaus ist die Visualisierung Parallele Koordinaten besser geeignet, da sich die Attributwerte schnell ablesen lassen und die Anwender intuitiv zurecht finden können.

3.3.3 Visualisierung Drei

Bei der Visualisierung Drei geht es um eine Visualisierungstechnik namens Stickfigureplot. Eine wichtige Eigenschaft dieser Technik besteht darin, dass sich mehrere Attributwerte miteinander vergleichen lassen. Dadurch können mehrere Zusammenhänge zwischen den Variablen untersucht werden, um wichtige Rückschlüsse aus den Datensätzen zu ziehen.

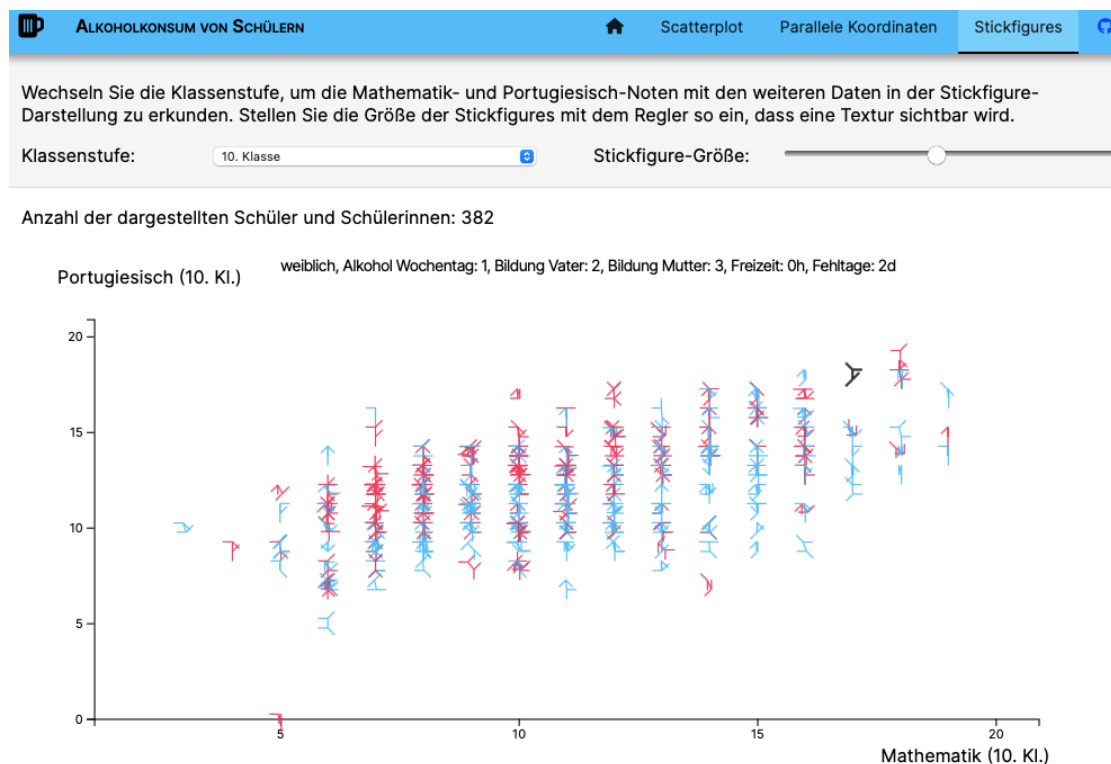


Abbildung 3: Stickfigureplot (Quelle: eigene Darstellung)

Aus der dargestellten Abbildung 3, lässt sich eine ausgewählte Stickfigur und deren Attributwerte deutlich erkennen. Wenn eine bestimmte Stickfigur ausgewählt wurde, wird diese automatisch farblich markiert. Somit kann erkannt werden welche Stickfigur ausgewählt wurde. Weiterhin können Korrelationsbeziehungen identifiziert werden, durch das Vertuschen der x- und y-Achen. Bei dieser Visualisierung werden im Vergleich zu den bisherigen Visualisierungstechniken die x- und y-Achse nicht unabhängig voneinander ausgewählt, sondern die Kombination der Achsen ist hierbei bereits vorgegeben. Diese Entscheidung wird so begründet, dass im Vorfeld viele Alternativen durchgeführt wurden und auf ihre Sinnhaftigkeit mehrfach geprüft wurden, sodass es sich bei diesen Kombinationen bereits um sinnvolle Verknüpfungen handelt.

Alternativ zu dieser Visualisierung kann eine Graph-Darstellung namens Sugiyama-Methode dargestellt werden. Allerdings würde diese Darstellung aufgrund der große Anzahl der vorhandenen Datensätzen unübersichtlich sein und zu falsche interpretation der Datensätze führen.

3.4 Interaktion

Bei den drei vorgestellten Visualisierungstechniken können viele Interaktionsmöglichkeiten durchgeführt werden. Die Visualisierungstechniken sind so aufgebaut, damit möglichst viele Interaktion zwischen den Variablen entstehen. Aus der Interaktion der Attributwerte, lassen sich wichtige Erkenntnisse aus den Daten extrahieren. Bei den Visualisierungstechniken können viele Interaktionsmöglichkeiten durch beliebiges auswählen der x- und y-Achse entstehen. Im Gegensatz dazu sind die Interaktionsmöglichkeiten der Stickfigureplot etwas reduzierter, da der Austausch der Achsen vordefiniert ist. Dabei soll betont werden, dass durch das Gegenüberstellen der sieben Attribute die Stickfigureplot möglichst interessante und wichtige Zusammenhänge entstehen können. Daten visuell und interaktiv darzustellen, gibt den Zielgruppen bessere Möglichkeiten, gewisse Muster und Trends zu erkennen. Die Zielgruppen können individuell festlegen, aus welcher Darstellung der Daten sie welche Erkenntnisse ziehen. Viele Attribute aus den Datensätzen weisen wichtige Kennzahlen auf, die besonders für die Sozialforschung interessant sind, um Hypothesen aus der Darstellungen zu generieren. Die Anwender können durch Beobachtung bestimmte Muster beziehungsweise Korrelationsbeziehungen der Datensätze neue Einblicke gewinnen, um beispielsweise Handlungsempfehlungen aus den Visualisierungstechniken zu extrahieren. Aus der Visualisierungen von Scatterplot und Stickfigureplot können zudem erkannt werden, ob ein linearer Zusammenhang zwischen zwei Variablen besteht. Darüberhinaus können Aussagen über bestimmten Merkmale getroffen werden. Zwischen den Visualisierungstechniken Stickfigureplot und Scatterplot wurde Interaktion zwischen den ausgewählten Datensätze aufgebaut. Dadurch wird

folgendes Interaktion entstehen: wenn aus dem Stickfigureplot raus gewechselt wird, werden die gewählten Daten aus dem Stickfigureplot für den Scatterplot updated. Es ist anzumerken, dass der Interaktion so aufgebaut ist, dass man nur durch Anklicken eines Datensatzes aus der Visualisierung Stickfigureplot die Daten für den Scatterplot updated werden aber nicht umgekehrt.

Zu den unterschiedlichen Visualisierungstechniken wird eine Webseite aufgebaut, die alle ausgewählten Visualisierungen der Daten umfasst. Die Anwender können jederzeit die entsprechenden Visualisierungen abrufen. Durch Anklicken sowohl der Navigationsleiste als auch der dargestellten Buttons können die Anwender die Visualisierungstechniken Scatterplot, Parallele Koordinaten und Stickfigureplot ansehen und beliebig zwischen den Visualisierungen hin und her wechseln. Die Webseite für die verschiedenen Visualisierungstechniken wurde mithilfe von GitHub Pages aufgebaut.

4. Implementierung

Für die Implementierung der Visualisierungstechniken waren die erworbenen Kenntnisse in den Übungen und Vorlesungen des Moduls “Information Retrieval und Visualisierung“ sehr hilfreich. Hinsichtlich der Programmierung des Scatterplots wurden die Inhalte aus den Übungen eins bis drei verwendet. Zur Erstellung der Parallele Koordinaten wurde als Grundlage für den Code die Lerninhalte aus der Übung sechs und sieben genommen. Zudem wurde auch eine Literaturrecherche durchgeführt, die als Hilfestellung genutzt wurde, um bestimmte Funktionen in der Programmiersprache Elm zu definieren. Der Inhalt des Codes der dritten Visualisierungstechnik Stickfigur basiert auf den Lerninhalten der Übung acht und der im StudIP zur Verfügung gestellten Lerninhalte. Die Programmierinhalte der drei dargestellten Visualisierungstechniken werden so strukturiert, dass die Anwender eine Übersicht zu den geschriebenen Codes erhalten. In vielen Abschnitten in der Programmierung sind Kommentare hinzugefügt, damit der Code für Außenstehende gut lesbar und nachvollziehbar ist.

Für den Aufbau der Visualisierungen wurden verschiedene Module aus der Webseite elm-lang.org lokal auf dem Rechner installiert. Diese Module werden zu dem Projekt im Visual Studio Code (VSC) “Student Alcohol Consumption” hinzugefügt um bestimmte Parameter und Funktionen zwecks der Implementierung zu verwenden. In diesem Abschnitt werden die wichtigsten Funktionen zusammengefasst, die fundamental für die Visualisierung der Daten in Elm sind. Für die drei Techniken wurde eine Datei namens Main gebaut, in welcher Funktionen der Browser.element: init für die Initialisierung, die Sektion view, subscriptions, und update definiert wurden. Diese

Funktionen sind innerhalb der Main zu finden. Die Interaktionsmöglichkeiten sind in Main dynamisch aufgebaut und Anwender können anhand der dargestellten Buttons und Navigationsleiste auf die entsprechenden Visualisierungen gelangen.

Die Buttons dienen dazu zwischen den Merkmalen hin und her zu springen um möglichst viele Zusammenhänge in den Daten zu erkunden.

Zusammengefasst ist Main für das Switchen von den Visualisierungen und Bündeln von Nachrichten, die aus der jeweiligen Visualisierungen gesendet werden, verantwortlich. Dieses Bündeln erfolgt durch das Mapping und das Wrapping jedes Models, jeder Message, und jedes Views der drei Visualisierungen.

Für das Laden der Datensätze in Elm wird der Link von Github genommen, in welchem die CSV-Daten hochgeladen sind. Die Decodierung der entnommenen Datensätzen aus Github wird mithilfe der CSV-Decoders, als “decodingStudentAcoholConsumption” definiert, erfolgen. Weiterhin wird eine weitere Funktion update benutzt, damit die Anwender die x- und y-Achsen beliebig austauschen können. Für die Visualisierung der Darstellungen wurden zwei Techniken verwendet: SVG und CSV.

Die bereits erwähnten Funktionen, entnommen aus den Übungsinhalten, mussten bei der Programmierung der Visualisierungstechniken angepasst werden. Dadurch dass mehrere Attribute zur Analyse herangezogen wurden, mussten Anpassungen der Übungsinhalten für den Projektbericht vorgenommen werden.

Notwendige Parameter wurden definiert, um Interaktion zwischen den Visualisierungen zu ermöglichen. In jeder Visualisierung wurde der Parameter Chosen Data, der innerhalb type alias Data gehört, definiert, das den type Maybe Studentalcoholconsumption hat. Der Parameter Chosen Data wurde für jeden View, Update und Message definiert um bestimmte Ereignisse zu empfangen. Weiterhin wurde der Parameter PointChosen, der innerhalb type MSG zu finden ist, definiert um einen bestimmten Datensatz aus der Visualisierung auszuwählen. Die Parameter sind so definiert, damit Interaktion zwischen den Visualisierungen entstehen kann. Die definierten Parametern und Funktionen ermöglichen das bündeln den Nachrichten zwischen den Visualisierungen und werden durch das Main gesteuert. Die Anwender haben die Möglichkeit bestimmte Datenpunkte durch Mausklick von der Visualisierung Stickfigureplot auszuwählen. Wenn aus dem Stickfigureplot raus gewechselt wird, werden die gewählten Daten aus dem Stickfigureplot für den Scatterplot updated. Darüber hinaus wurden in der Datei Main mögliche Anpassungen innerhalb Update sowohl für Stickfigureplot als auch Scatterplot vorgenommen. Abbildungen 4 und 5 stellen die Interaktion zwischen der zwei Visualisierungen dar.

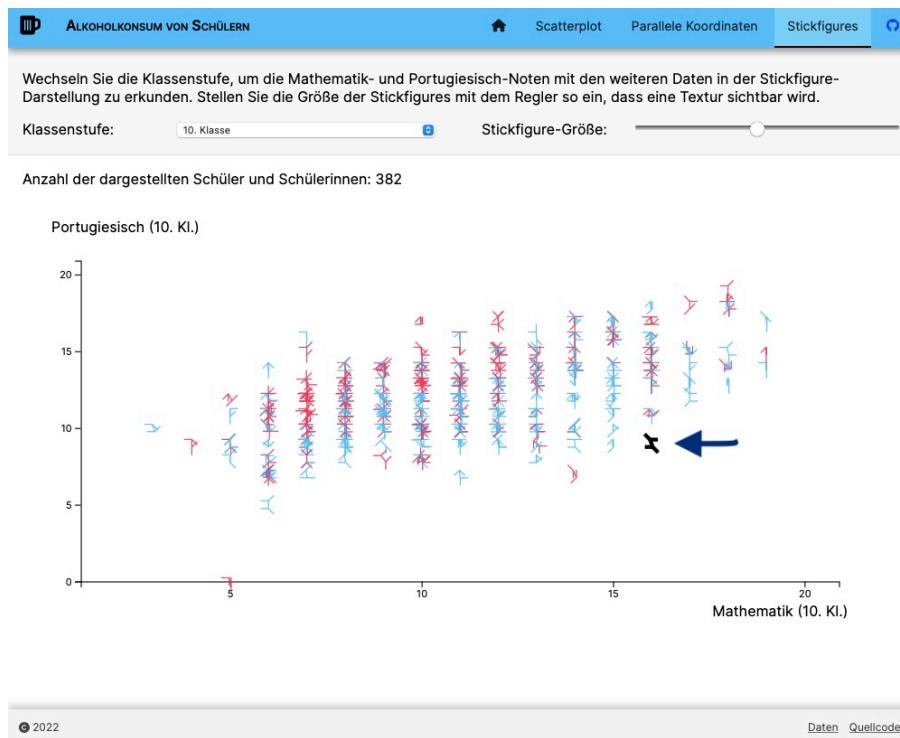


Abbildung 4: Stickfigureplot (Quelle: eigene Darstellung)

Die Interaktion ist dynamisch aufgebaut und die Anwender können individuell entscheiden, welche Interaktionsmöglichkeit zwischen der Stickfigureplot und Scatterplot sie gerne untersuchen möchten. Es ist dabei anzumerken, dass nur eine Interaktion zwischen der Visualisierungen Stickfigureplot und Scatterplot möglich ist.

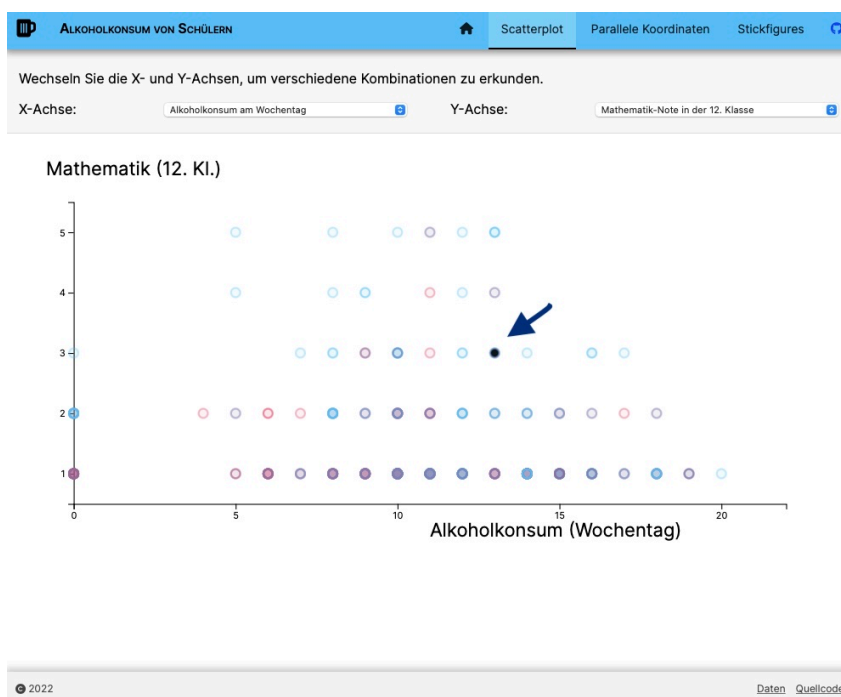


Abbildung 5: Scatterplot (Quelle: eigene Darstellung)

Der Code für alle drei Visualisierungstechniken befindet sich in einem Unterordner mit dem Name src aus dem Gesamtordner GitHub. Als Grundlage für CSS, also das Design der Visualisierungen und Html u.a. dargestellten Button, Dropdown und Navigationsleiste, wurden neben der Inhalte auf der Webseite Elm, eine Literaturrecherche durchgeführt und oft wurden Beispiele in unterschiedlichen GitHub-Projekten die für das Visualisierungsprojekt hilfreich sind, gefunden.

5. Anwendungsfälle

Um den Anwendern mit den vorgestellten Visualisierungen vertraut zu machen, werden in diesem Kapitel Anwendungsbeispiele beschreiben, die für entsprechenden Zielgruppen mögliche Anwendungsfälle darstellen.

5.1 Anwendung Visualisierung Eins

Bei der Visualisierung Eins wird der Output der y-Achse, welche für die Abschlussnote aus dem Schulfach Mathematik steht in Verhältnis zu der x-Achse, welche für den Alkoholkonsum unter der Woche steht dargestellt. Sobald ein Kreis ausgewählt wird, werden die entsprechenden Attributwerte angezeigt. Betrachtet man die Datensätze als Ganzes kann man erkennen im welchen Notenbereich die Schüler liegen, in Abhängigkeit des Ausmaßes des Alkoholkonsums. Zudem kann man aus den Daten erkennen, ob sich mögliche Ausreißpunkte aus den Daten identifizieren lassen. Die Abbildung 6 dient der Zielgruppen Schuldirektoren, Sozialforschung und Data Science, um beispielsweise die Frage zu beantworten, wie der Alkoholkonsum unter der Woche die Abschlussnoten in dem Studienfach Mathematik beeinflusst. Wie in der Abbildung zu sehen ist, liegen die meisten Datenpunkte zwischen den Alkoholstufen eins und zwei und die erreichten

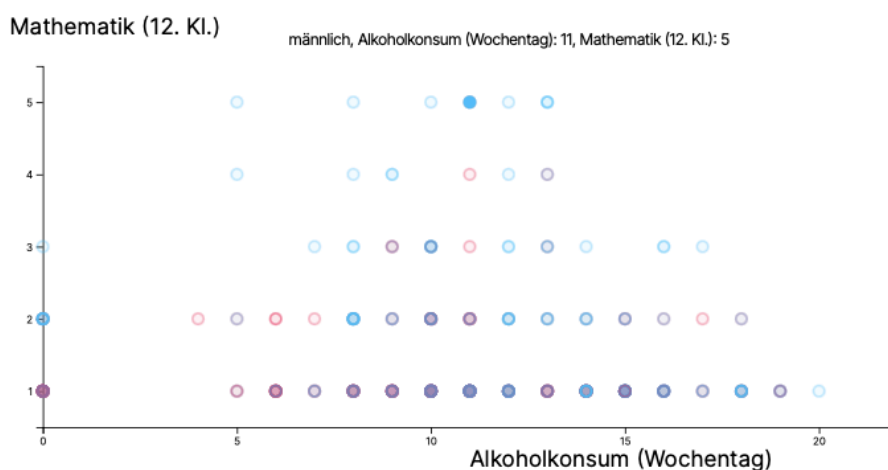


Abbildung 6: Scatterplot (Quelle: eigene Darstellung)

Abschlussnoten zwischen 15 und 17 Punkten. In welchem Bereich die meisten Schüler liegen, ist farblich gekennzeichnet. Die Betonung der Farbe des Außenkreises der dargestellten Datenpunkte, ist ein Hinweis dafür, dass mehrere Datensätze übereinander liegen. In der folgenden Abbildung kann auch erkannt werden, dass bei einigen Kreisen im oberen rechten Bereich die Farbe heller ist was bedeutet, dass nur wenig Schüler die Alkoholstufe fünf erreichen und im Vergleich zu den anderen Schülern, deren Alkoholkonsum in dem Bereich eins bis zwei liegt, schlechtere Noten bekommen.

In der Scatterplot kann durch eine Stichprobe herausgefunden werden, welche Geschlechter die höchsten Alkoholstufen erreichen, indem die Anwender durch das Anklicken der dargestellten Kreise, die im Bereich der Alkoholstufe fünf liegen, das Geschlecht der Schülern in Erfahrung bringen. Darüberhinaus kann ein Schuldirektor erkennen, welche Alkoholstufe die meisten Schüler erreichen und wie die Abschlussnote in Zusammenhang mit dem Alkoholkonsum aussieht. Die Sozialforschung kann aus den dargestellten Datenpunkte den Zusammenhang der Abschlussnoten mit dem höchsten Alkoholkonsum unter der Woche untersuchen. Wie in der Abbildung zu sehen ist erreicht nur ein kleiner Teil der Schüler die Alkoholstufe fünf und die Noten schwanken zwischen 10 und 20. Data Science-Unternehmen kann zum Beispiel dafür interessieren, ob Ausreißer in der Visualisierungen zu finden sind die Schülerleistungen beeinflussen können oder welche Aussagen allgemein über die Daten gemacht werden können.

5.2 Anwendung Visualisierung Zwei

Durch Analyse der Visualisierungstechnik Parallele Koordinaten, können vier Merkmale in Betracht gezogen werden und interessante Erkenntnisse gewonnen werden.

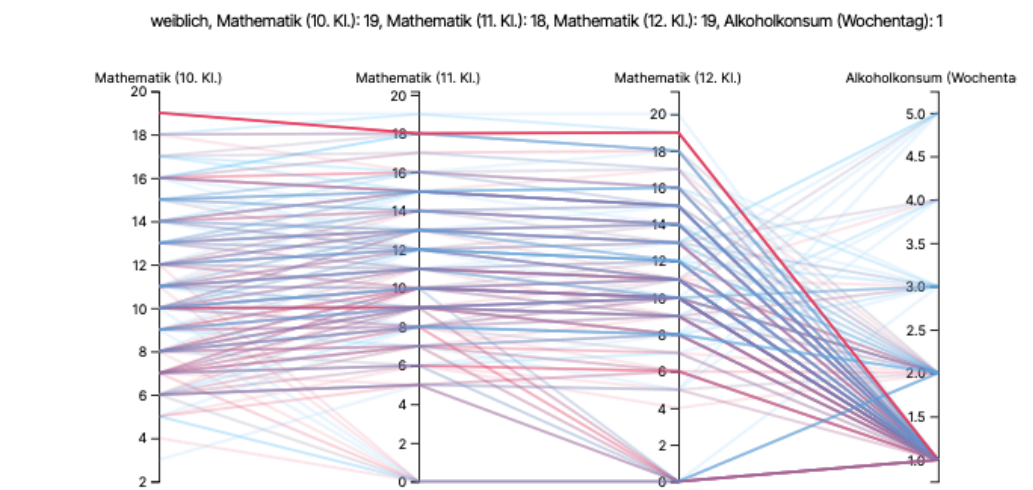


Abbildung 7: Parallele Koordinaten (Quelle: eigene Darstellung)

In der Abbildung sind die Schulnoten von zehnter, elfter und zwölfter Klasse dargestellt in Zusammenhang mit dem Alkoholkonsum unter der Woche.

Die Anwender können aus der dargestellten Technik beobachten, wie die erreichten Schulnoten aus der zehnten und elften Klasse in Zusammenhang mit dem Alkoholkonsum in die Gesamtnote mit einfließen. Die Daten können einzeln und verbunden betrachtet werden.

Der Schuldirektor aus der vordefinierten Zielgruppe kann sich für ein Gesamtbild hinsichtlich der Ergebnisse der Schüler interessieren. Er kann die Entwicklung der Schulnoten aus der Alkoholstufe 1 und 2 beobachten. Auf der Abbildung kann auf den ersten Blick erkannt werden, dass die Schüler bis zur Alkoholstufe 1, die besseren Schulnoten für die zehnte, elfte und zwölfe Klasse erreichen.

So kann auch festgestellt werden, dass diejenigen, die in der zehnten und elften Klasse sehr gute Noten erreichen und in dem Bereich der Alkoholstufe 1 bleiben, somit die besten Noten haben. Aus der dargestellten Visualisierungstechnik kann zudem auch erkannt werden, dass die Schüler, die in der zehnten und elften Klasse in dem Bereich von 0 bis 9 liegen, auch diejenigen sind, die einen Schulabschluss nicht schaffen. Dadurch können die Schuldirektoren zukünftig für diejenigen, die in der zehnten Klasse schlecht abschneiden zusätzlich Schulunterricht in der Schule anbieten, damit die Schüler in der kommenden elften und zwölften Klasse die Möglichkeit erhalten die Schulnoten zu verbessern. Für die Sozialforschung wäre zudem auch interessant zu untersuchen, ob die Schulnoten von der zehnten, elften und zwölften Klasse stark voneinander abhängen. Aus den dargestellten Daten können Stichproben gezogen werden um herauszufinden, ob beispielsweise die Schüler, welche eine schlechte Note in der zehnten Klasse bekommen, sich in den folgenden Klassen verbessert haben oder gleich geblieben sind. Für die Zielgruppe Data Science sind ebenfalls die Untersuchungen der Zusammenhänge der Variablen von Interesse.

So können bestimmte Gruppen von Schülern ausgehend von erreichten Schulnoten in der zehnten, elften und zwölften Klasse in Abhängigkeit mit der Alkoholstufe untersucht werden.

Wie in der Abbildung 7 zu sehen ist, werden die besten Abschlussnoten von den früheren Noten aus der zehnten und elften Klasse beeinflusst und das erreichte Trinkniveau beträgt 1. Die höchste Abschlussnote, die Schüler in der 5. Alkoholstufe erreicht haben beträgt 13. Aus den Daten kann erkannt werden, dass mit der Erhöhung der Alkoholstufe die Abschlussnoten deutlich schlechter ausfallen. So können die gewonnenen Erkenntnisse aus den Daten öffentlich zugänglich gemacht und für spätere Reports genutzt werden. Reports können wiederum für die Wissenschaft sehr hilfreich sein damit mögliche Abweichungen in den untersuchten Variablen aufgedeckt werden.

5.3 Anwendung Visualisierung Drei

Die Visualisierung Drei Stickfigureplot versucht die Frage zu beantworten, welchen Einfluss die Bildung der Eltern in Kombination mit Alkoholkonsum unter der Woche, Freizeit und Abwesenheit auf die Abschlussnote hat. Den Zusammenhang zwischen diesen Variablen zu untersuchen ist sehr wichtig, da die Abschlussnote ein Entscheidungskriterium für eine Weiterführung in Form eines Studiums an einer Universität ist. Wie in der dargestellten Abbildung 8 zu sehen ist, lassen sich sieben einzelne Variablen der jeweiligen Schüler miteinander vergleichen.

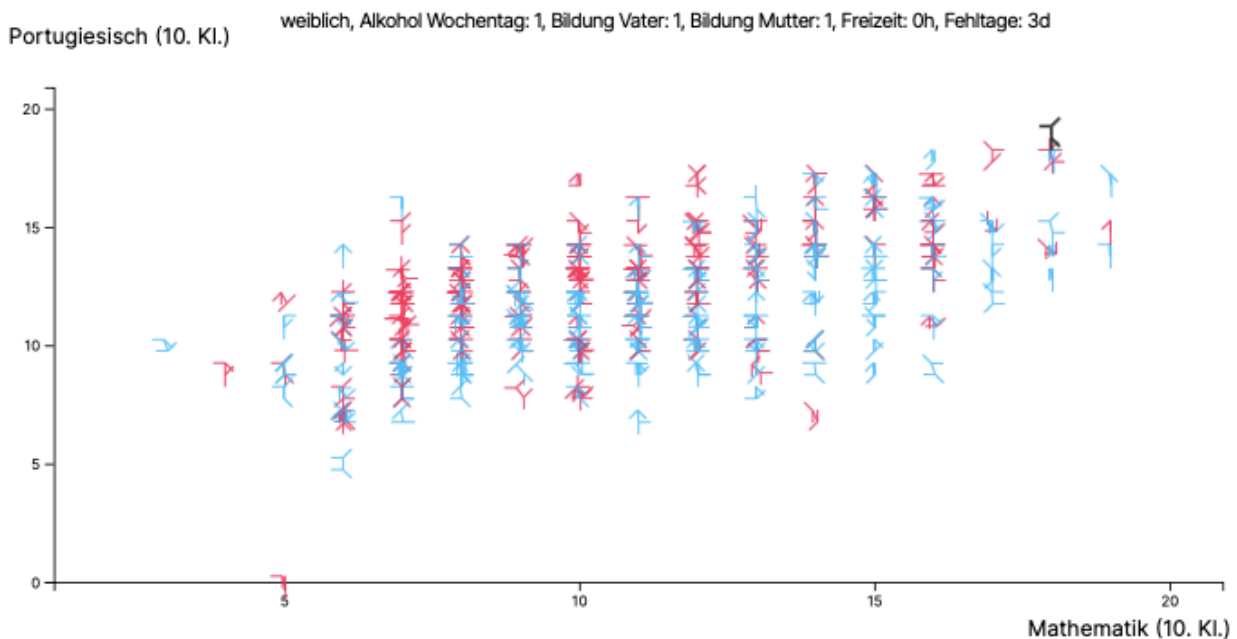


Abbildung 8: Stickfigureplot (Quelle: eigene Darstellung)

Zudem können die Daten als Ganzes betrachtet werden um zu sehen welche Korrelationsbeziehungen sich aus der dargestellten Visualisierung ableiten lassen. Darüberhinaus lässt sich aus der Abbildung 8 feststellen, dass die erreichten Schulnoten für Mathematik und Portugiesisch aus der zehnten Klasse mit den anderen dargestellten Variablen (Alkoholkonsum, Freizeit nach der Schule, Abwesenheit und die Bildung der Eltern) positiv korrelieren. Mithilfe dieser Visualisierung lassen sich korrelative Beziehungen identifizieren. In diesem Fall kann aus der dargestellten Abbildung erkannt werden, dass die Datensätze einen positiv linearen Zusammenhang darstellen. Somit kann die Zielgruppe Data Science Erkenntnisse über Korrelationsbeziehungen der Daten gewinnen. Stichprobenartig kann durch das Anklicken der jeweiligen Stickfiguren festgestellt werden, dass diejenigen Schüler, die öfter in der Schule fehlen auch viel Alkohol konsumieren, sich nach der Schule viel Freizeit nehmen und deren Eltern keinen Akademischen Abschluss haben, schlechte Schulnoten in beiden Studienfächer erreichen.. Zudem gibt es auch Schüler, die trotz des

niedrigen Bildungsniveau der Eltern gute Note in der Schule erzielen. Jedoch kann aus dieser Visualisierung nicht auf den ersten Blick erkannt werden, ob das Bildungsniveau der Eltern ein Problem für die Leistung darstellt. Diese Analyse der Daten kann für die Zielgruppe der Schuldirektoren nützlich sein, damit zukünftig Methoden in der Lehre eingeführt werden, um Schüler, die nicht aus akademischer Familie stammen, in der Schule zu unterstützen. Zudem könnten diese Erkenntnisse der Sozialforschung dabei helfen, die Fragen zu beantworten, wie und warum die demographischen, sozialen und schulbezogenen Merkmale die Schülerleistungen beeinflussen.

6. Verwandte Arbeiten

Einige vergleichbare wissenschaftliche Visualisierungen und Berichte werden in diesem Kapitel beschrieben. Ein Artikel mit dem Titel “Visualisation and optimisation of alcohol-related hospital admissions ICD-10 codes in Welsh e-cohort data” verwendet anders als dieser Arbeit zwei verschiedene Visualisierungstechniken Säulendiagramm und Akkorddiagramm um den Code der Diagnose bei Krankenhauseinweisung und -entlassung miteinander visuell zu vergleichen und den Code der Diagnose mithilfe der verwendeten Visualisierungstechniken zu ermitteln, die für eine alkoholbedingten Krankenhauseinweisung notwendig ist.¹³ Der Akkorddiagramm wird wie folgt dargestellt.

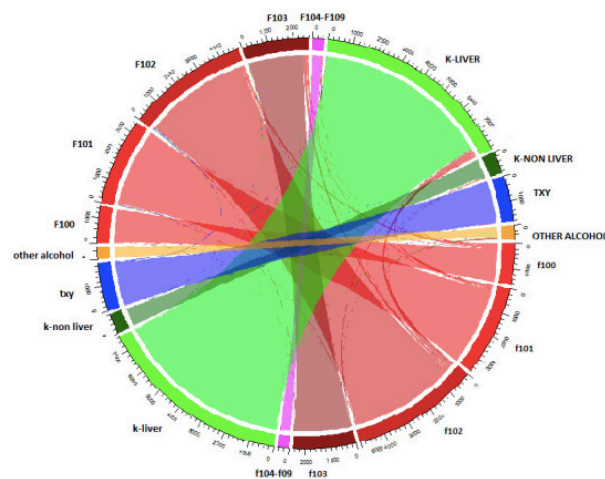


Abbildung 9: Akkorddiagramm der ersten Alkoholcodes für alkoholbedingte Einweisungen und Entlassungen¹⁴

¹³ Vgl. Trefan L., Akbari A., Siân M. S., Farewell M. D., Fone D., Lyons A. R., Hywel J. M. & Moore C. S. (2021)

¹⁴ Trefan L., Akbari A., Siân M. S., Farewell M. D., Fone D., Lyons A. R., Hywel J. M. & Moore C. S. (2021)

Eine weitere wissenschaftliche Artikel verwendet den Regression Lens Methode um globalen und lokalen Regressionsmodellen in Streudiagrammen Daten zu untersuchen. Eine Abbildung wird wie folgt dargestellt.¹⁵

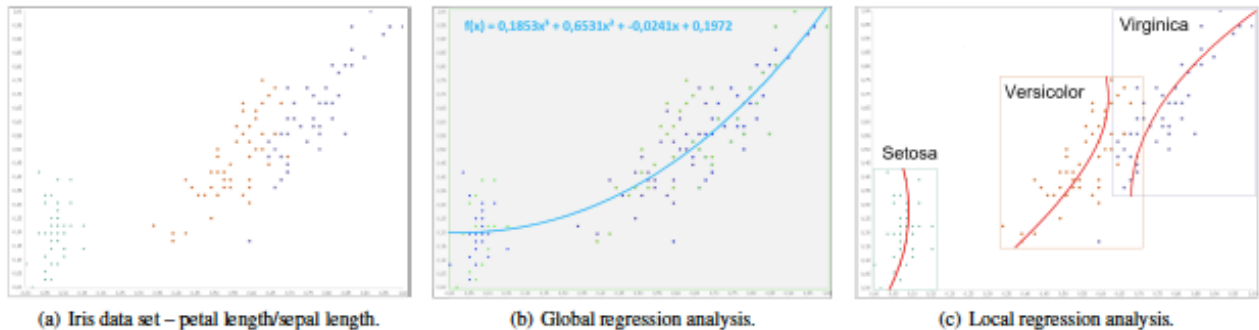


Abbildung 10: Darstellung Iris-Datensatz und globaler & lokaler Regressionsanalyse¹⁶

Eine weitere wissenschaftliche Artikel stellt mithilfe einer Scatterplot die Visualisierung der Finanzdatensätzen mit den Anleihekursen die zwei Jahre umfassen dar.¹⁷

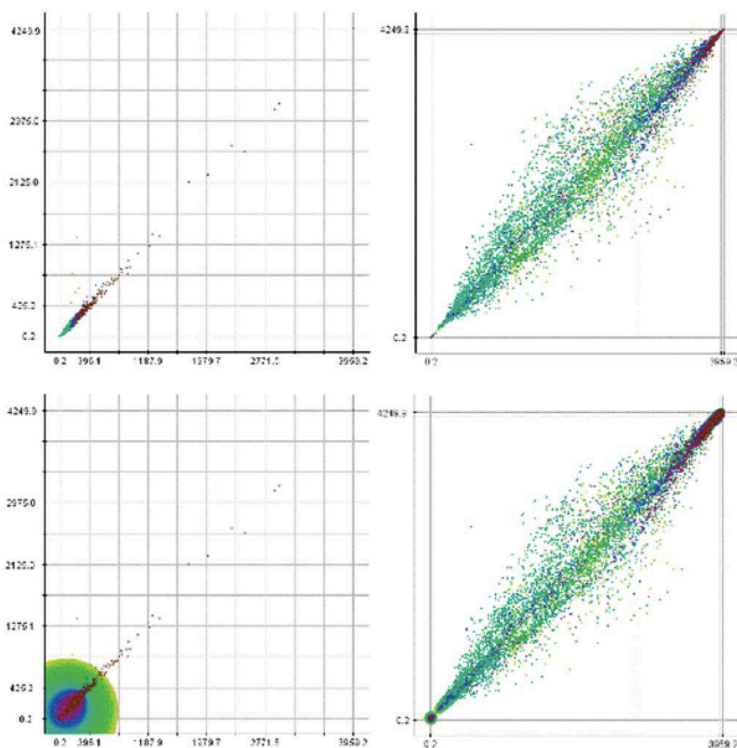


Abbildung 10: Darstellung Financial data set showing bond market prices of two consecutive years.¹⁸

¹⁵ Vgl. Shao L., Mahajan A., Schreck T., & Lehmann J. D. (2017)

¹⁶ Vgl. Shao L., Mahajan A., Schreck T., & Lehmann J. D. (2017)

¹⁷ Vgl. Trefan L., Akbari A., Siân M. S., Farewell M. D., Fone D., Lyons A. R., Hywel J. M. & Moore C. S. (2021)

¹⁸ Trefan L., Akbari A., Siân M. S., Farewell M. D., Fone D., Lyons A. R., Hywel J. M. & Moore C. S. (2021)

Neben den vorgestellten Studien bzw. Visualisierungsprojekten wurden drei andere Studien gefunden, in denen einige statistische Methoden verwendet wurden, die den Zusammenhang des Alkoholkonsums mit dem Lerneffekt in der Schule untersuchen. Darüberhinaus wird in den Studien berichtet, dass der Alkoholkonsum nicht immer auf die Beeinträchtigung der Schulnoten hindeutet. Der Lerneffekt in der Schule wird oft von dem Ausmaß an Alkohol beeinflusst und die Schülerleistung kann geschlechtsspezifisch unterschiedlich ausfallen.¹⁹

7. Zusammenfassung und Ausblick

Mithilfe der dargestellten Visualisierungstechniken erhalten die Anwender einen umfangreichen Überblick zum Thema vom Einfluss des Alkoholkonsums auf die Schulnoten in Portugal. Zudem ermöglichen die Techniken Interaktive Möglichkeiten ausgewählter Attributwerte und liefern interessante Erkenntnisse bezüglich der Untersuchung vom Einfluss des Alkoholkonsums und demographischen und sozialen Merkmalen auf die Schulnoten.

Hinsichtlich der vorgestellten Visualisierung Scatterplot können Informationen über die ausgewählten Attribute angezeigt werden und zudem lässt sich anhand der ausgewählten Farben eindeutig identifizieren, um welches Geschlecht es sich handelt. Weiterhin liefert diese Technik Erkenntnisse über die Korrelationsbeziehungen der visualisierten Datensätze. Die wesentliche Eigenschaft dieser Visualisierungstechnik ist der Vergleich zweier Attributwerte in einem zweidimensionalen Raum. Diese Technik ist so aufgebaut, dass die Anwender die Auswahl der x- und y-Achsen beliebig wählen können. Durch diese Interaktionsmöglichkeit, lassen sich schnell und intuitiv Vergleiche zweier Werte machen. Anhand der Visualisierungstechnik Parallele Koordinaten können gleichzeitig vier Attributwerte miteinander verglichen werden. Bei dieser Technik können die Anwender entscheiden, welche Interaktionsmöglichkeiten sie untersuchen möchten. Zum Beispiel kann mithilfe dieser Technik, wie bereits im Kapitel 5 erwähnt, festgestellt werden, ob ein Einfluss von dem Alkoholkonsum auf die Schulnoten ausgehend von der zehnten bis elften Klasse der Sekundarschule auf die Abschlussnoten besteht. In Kapitel 5 wurde ebenso festgestellt, dass die Schulnoten aus der ersten und zweiten Klasse, die Abschlussnoten stark beeinflussen. Es ist anzumerken, dass mithilfe dieser Technik keine Voraussage der Schülerleistungen getroffen werden kann.

Die dritte Visualisierungstechnik Stickfigureplot stellt die Datensätze in einem mehrdimensionalen Raum dar. Darüberhinaus werden in dieser Darstellung neben schulbezogener Leistung, soziale und

¹⁹ Vgl. Balsa, Giuliano & French (2011), DeSimone & Wolaver (2005), Dee & Evans (2003)

demographische Merkmale in Erwägung gezogen. Die Anwender können Einblicke darüber erhalten, wie die sozialen und demographischen Faktoren in Zusammenhang mit der Abschlussnoten stehen.

Zu den verwendeten Visualisierungstechniken in dieser Projektarbeit können auch mögliche Anpassungen vorgenommen werden. Denkbar ist es, die Daten für Schüler getrennt zu visualisieren, also in männliche und weibliche Schüler. Weiterhin kann auch neben Buttons, Dropdown- und Navbar-Menü, zusätzlich ein Drag and Drop oder auch Flyout-Menü dargestellt werden. Eine Mischung der Elemente - Navbar-, Flyout-Menü und Drag and Drop - zur Zuweisung des Attributs auf der x- und y-Achse ist auch eine interessante Darstellung. Bei der Stickfigureplot-Darstellung ist es denkbar, die Stickfiguren so darzustellen, dass, sobald die Anwender eine bestimmte Stickfigur auswählen, die anderen Stickfiguren aus der Visualisierung ausgeblendet werden. Somit werden nur die ausgewählten Stickfigur im Vordergrund stehen und die zugehörigen Attributwerte werden zu den einzelnen Bestandteile der Stickfigur positioniert.

Eine Überlegung ist auch, andere Variablen in den vorgestellten Visualisierungen in Erwägung zu ziehen, zum Beispiel demographische Merkmale wie die Zeit, die die Schüler von zu Hause bis zu der Schule benötigen.

Litterature:

Joos Th., & Schmitz P. (2021): Sicherheit in Oracle-Datenbanken konfigurieren. Security-insider: Vogel Communications Group. URL: <https://www.security-insider.de/sicherheit-in-oracle-datenbanken-konfigurieren-a-1031255/>, Abruf am 10.03.2022.

P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUTURE BUSINESS TECHNOLOGY Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7.

Koler P., (2014): Rausch und Identität Jugendliche in Alkoholszenen. 1. Auflage. URL: https://www.forum-p.it/smartedit/documents/inhaltelements/_published/Rausch&Identität.pdf, Abruf am 15.08.2022.

U.S. Department of Health & Human Services (2016): Why Do Adolescents Drink, What Are the Risks, and How Can Underage Drinking Be Prevented? URL: <https://pubs.niaaa.nih.gov/publications/AA67/AA67.pdf>, Abruf am 07.09.2022.

Notenumrechnung Portugal. URL: https://www.uni-potsdam.de/fileadmin/projects/international/docs/Notenumrechnung_Länder/Notenumrechnung_Portugal_IO.pdf, Abruf am 09.09.2022.

Doerfel A. Helmholtz-Gemeinschaft (2018): Wenn das Gehirn trunken ist. URL: <https://www.helmholtz.de/newsroom/artikel/wenn-das-gehirn-trunken-ist/>, Abruf am 09.09.2022.

Keim A. D., Hao C. M., Dayal U., Janetzko H. & Baka P. (2019): Generalized scatter plots. URL: <https://journals.sagepub.com/doi/epdf/10.1057/ivs.2009.34>, Abruf am 20.12.2022.

Yi M. A Complete Guide to Scatter Plots. URL: <https://chartio.com/learn/charts/what-is-a-scatter-plot/>, Abruf am 09.09.2022.

Last Updated (2022): Decision Tree Algorithm Examples In Data Mining. URL: <https://www.softwaretestinghelp.com/decision-tree-algorithm-examples-data-mining/>, Abruf am 09.09.2022.

Trefan L., Akbari A., Siân M. S., Farewell M. D., Fone D., Lyons A. R., Hywel J. M. & Moore C. S. (2021): Visualisation and optimisation of alcohol-related hospital admissions ICD-10 codes in Welsh e-cohort data, Abruf am 20.12.2022.

Yanasagaran S., Tan J., Che J. & Adinda A. (2018): RPubs - R PROJECT 1 (ENVX1002/DATA1001) URL: <https://rpubs.com/Jonchee0903/377825>, Abruf am 09.09.2022.

Pickett M. R. & Grinstein G. G. (1988): Iconographic Displays For Visualizing Multidimensional Data. URL: <https://www.researchgate.net/publication/3767640>, Abruf am 09.09.2022.

Gemignani Z. (2021): Better Know a Visualization: Understanding Parallel Coordinates Charts. URL: <https://www.juiceanalytics.com/writing/writing/parallel-coordinates>, Abruf am 09.09.2022.

Balsa A. L., Giuliano L.M., & French M T. (2011): The effects of alcohol use on academic achievement in high school. NIH-PA Author Manuscript, Abruf am 09.09.2022.

DeSimone J. & Wolaver A. (2005): Drinking and Academic Performance in High School. Working Paper 11035 <http://www.nber.org/papers/w11035>, Abruf am 30.09.2022.

Dee Th. S. & Evans W. N. (2003): Teen Drinking and Educational Attainment: Evidence from Two-Sample Instrumental Variables Estimates. Journal of Labor Economics, vol. 21, no. 1, University of Chicago.

Soyka, M. (2001): Serie - Alkoholismus: Psychische und soziale Folgen chronischen Alkoholismus. URL: <https://www.aerzteblatt.de/archiv/29088/Serie-Alkoholismus-Psychische-und-soziale-Folgen-chronischen-Alkoholismus>, Abruf am 28.09.2022.

Gonçalves I. A., & de Sousa Carvalho A. A. (2017): Pattern of alcohol consumption by young people from North Eastern Portugal. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5836530/pdf/med-12-494.pdf>, Abruf am 30.08.2022.

Stumpp G., Stauber B., & Reinl H. (2009): Einflussfaktoren, Motivation und Anreize zum Rauschtrinken bei Jugendlichen. URL: https://www.berlin-suchtprevention.de/wp-content/uploads/2016/10/2009_Jugendliche_und_Rauschtrinken_BMG.pdf, Abruf am 15.08.2022.

FOCUS online, (2017): Studie behauptet: Wer Alkohol trinkt, schreibt danach die besseren Prüfungen. URL: https://www.focus.de/wissen/mensch/wunderlernmittel-fuer-studenten-studie-belegt-wer-alkohol-trinkt-schreibt-die-besseren-klausuren_id_7409530.html, Abruf am 15.08.2022.

Student Alcohol Consumption, update (2016). URL: <https://www.kaggle.com/datasets/uciml/student-alcohol-consumption?select=student-merge.R>, Abruf am 01.07.2022.

Shao L.,Mahajan A.,Schreck T., &. Lehmann J. D. (2017) Interactive Regression Lens for Exploring Scatter Plots. URL: https://www.researchgate.net/publication/317957265_Interactive_Regression_Lens_for_Exploring_Scatter_Plots, Abruf am 20.12.2022.

Anhang:

Git-Historie

- * 6b70461 (origin/main) (add Index.html, d%)
- | * dda3995 (HEAD -> main) (Merge branch 'main' of https://github.com/Ella199/Elm_Project-Student-Alcohol-Consumption, d%)
- * | e5bf562 (delete index.html, d%)
- * | f65517b (Add files via upload, d%)
- * | 69c42b8 (Update README.md, d%)
- * | 7ed0a26 (Update README.md, d%)
- | * d49f758 (add index.html, d%)
- | * bf55392 (update.DS_Store, d%)
- | * b398c65 (update Bericht, d%)
- | * ec5304c (Merge branch 'main' of https://github.com/Ella199/Elm_Project-Student-Alcohol-Consumption, d%)
- * | d489841 (Projektbericht, d%)
- | * cc65b68 (add Projektbericht, d%)
- | * a62505e (Merge branch 'main' of https://github.com/Ella199/Elm_Project-Student-Alcohol-Consumption, d%)
- * | 4735601 (Update README.md, d%)
- * | b12e1d3 (Update README.md, d%)
- * | c76b0f4 (Update README.md, d%)
- * | 40c4c41 (Update README.md, d%)
- * | aeBBBBda (Update README.md, d%)
- * | ba13f5e (Update README.md, d%)
- * | 504999a (Update README.md, d%)
- * | 9611c9b (Update README.md, d%)
- * | 2368b40 (Update README.md, d%)
- * | 0f93496 (Update README.md, d%)
- * | b782960 (Update README.md, d%)
- * | 7e407e9 (Update README.md, d%)
- * | 82bb322 (Update README.md, d%)
- | * d7871c6 (add some updates final, d%)

| * 585dff5 (Merge branch 'main' of https://github.com/Ella199/Elm_Project-Student-Alcohol-Consumption, d%)

* | 06a8bb8 (Update README.md, d%)

* | acf290a (Update README.md, d%)

* | e66aebf (Update README.md, d%)

* | 7d34856 (Update README.md, d%)

* | c7fc272 (Update README.md, d%)

* | e94aa5a (Readme_preFinal, d%)

| * 0a28568 (Update main Final, d%)

* f8aec5a (add some adaption FINAL, d%)

* 6c2f11a (Final, d%)

* 0c924a0 (add some adaption (Final), d%)

* 04c30a3 (Delete Parallelekoordinaten.html (old version), d%)

* bfb2f06 (Delete Scatterplot.html (old version), d%)

* b1e2925 (Delete Visualisationstickfigureplot.html (old version), d%)

* 6df98a4 (Delete Webpage.html (old version), d%)

* 3aae6e7 (add Main for interactions in Scatterplot und Stickfigureplot, d%)

* fa661c2 (add some adaption at Sticfigure-Visualisation FINAL, d%)

* 2e7215d (add some adaption at Scatterplot FINAL, d%)

* bfd1160 (adaption for interactions in visualisation Stickfigure, d%)

* a9cd862 (add some adaption for interaction in scatterplot, d%)

* d76bd1d (add some parameters and functions for interaktion, d%)

* c396ce4 (Add module Util for type sex, d%)

* 5badc51 (add some adaptation at Visualisationstickfigureplot, d%)

* 1053fa7 (remove class line from stickfigure, d%)

* 3edc780 (update pointLabel, d%)

* e952852 (add sexLabel and add sexFlag, d%)

* 8d27a3a (add some adaptation at update, d%)

* 8afcf13 (add type sex, d%)

* dae083f (remove button, d%)

* 8a26133 (add drop down, d%)

* 7e2b113 (add stylesheet, d%)

* 732f9cc (add some packages, d%)

- * bfe2566 (add adaptations, d%)
- * 31c31e4 (update code Parallele Koordinaten, d%)
- * 615888c (change Html text at error, d%)
- * 7d34a1b (change text at loading, d%)
- * ea11ac5 (update parameter success l, d%)
- * afe1506 (remove Html Button, d%)
- * b80ab34 (add sex Label, d%)
- * 339af67 (add html to change 4.Axis, d%)
- * a9616ae (add html to change 3. Axis, d%)
- * 6af58a1 (add html to change 2.Axis, d%)
- * 0ee6b32 (add html to change 1. Axis, d%)
- * 709c1a2 (add nav, d%)
- * b727646 (add change msg value to change axis, d%)
- * 42e361e (add stylesheet, d%)
- * 196dad7 (update svg, d%)
- * 395aca3 (add MoveAxisUp num to update, d%)
- * 499898b (update, d%)
- * 28c7376 (add some adaptation at updat, d%)
- * bb1f711 (add MoveAxisUp Int to type Msg, d%)
- * a40b657 (add type sex, d%)
- * 14c6290 (add type alias Data, d%)
- * a125617 (add some packages to ParalleleKoordinaten, d%)
- * 34fc7ec (add some updates in Scatterplot, d%)
- * c4006c7 (update, d%)
- * 02db1b9 (update, d%)
- * 57409aa (update view, d%)
- * 523a944 (update view, d%)
- * 1e95c3d (update view loading, d%)
- * bf5ac2b (remove Html Buttons, d%)
- * 315541d (add Html to change y-axis, d%)
- * 876b993 (add HTML to change x-axis, d%)
- * ecb2b9d (add drop down-Menu, d%)
- * 7cb2847 (add some adaptation at SVG, d%)

- * 4abba26 (add case yxPoint.sex for male and female, d%)
- * b3d5671 (add some adaptation at point, d%)
- * 3111463 (add sexFlag, d%)
- * 02d06da (add sexLabel, d%)
- * a41b152 (add some other adaptation, d%)
- * 65092ea (add type Sex, d%)
- * bfac64a (add type alias data and and other adaptation, d%)
- * fcd2be0 (add packages, d%)
- * 48b335b (update Main, d%)
- * 545ef89 (add some package, d%)
- * eba1817 (update View, d%)
- * 0c9c2a0 (remove old view, d%)
- * 054ad68 (add link from Github to Webpage, d%)
- * e5fbf21 (add link to Kaggle Dataset, d%)
- * 50833ce (add footer model, d%)
- * f8bfcbe (add Html.map VisualisationStickfigureMsg, d%)
- * 3bcc720 (add Html.map ParalleleKoordinatenMsg, d%)
- * 3b8621c (add Html.map ScatterplotMsg, d%)
- * 2b47d8e (add Html Button for VisualisationStickfigure, d%)
- * 4a42ef2 (add Html Button for ParalleleKoordinaten, d%)
- * ba507f1 (add Html button for scatterplot, d%)
- * 8cd08a6 (add Html for webpage, d%)
- * a411e42 (update nav model, d%)
- * d19584d (update design, d%)
- * 39e1e47 (add navbar, d%)
- * 9888763 (update Html, d%)
- * 8300aa7 (add Html style, d%)
- * 0ed41a3 (add update msg model VisualisationStickfigure, d%)
- * c082398 (add update msg model for Scatterplot and ParalleleKoordinaten, d%)
- * 2b7feb9 (add Html.map VisualisationStickfigureMsg, d%)
- * de546de (add Html Text, d%)
- * 2669972 (add Html Button for VisualisationStickfigure, d%)
- * bc209bd (add subscriptions, d%)

- * bec8783 (add some adaptation in init, d%)
- * 81eff48 (remove Browser.sandbox, add Browser.element, d%)
- * 35befd3 (import Scatterplot exposing (scatterplot) for main, d%)
- * b16174c (add Main and other adaptation, d%)
- * f200b45 (Adjustments on the homepage, d%)
- * e9bedc9 (add ParalleleKoordinaten to Html, d%)
- * c4d7b1a (add StickfigurePlot.elm to Html, d%)
- * 0c6cd08 (add elm make for Scatterplot.elm, d%)
- * 32d32a7 (add link to Stickfigure, d%)
- * 2ed9e17 (add link to Parallelekoordinaten, d%)
- * 4aba044 (add link to Scatterplot, d%)
- * b3c2638 (update Html Webpage, d%)
- * 6d0dadd (Add title Webpage, d%)
- * 2e7738e (add Html-Webpage (Homepage), d%)
- * 10ddc55 (remove some variables from Stickfigureplot, d%)
- * bf1ee56 (change the label of the y axis, d%)
- * aea78fa (change the label of the x axis, d%)
- * c30fc36 (update stickfigureplot, d%)
- * 84b59a3 (Add ChangeGrade g at update, d%)
- * a1c2d13 (add ChangeGrade (String) at type Msg, d%)
- * e190620 (Add gr: String in model, d%)
- * c7ad237 (update Model, d%)
- * 0e2a60c (update Code for html Button, d%)
- * ab1ecd8 (update Button still doest work, d%)
- * 1e2b87f (Button doest work, d%)
- * 2c520b8 (update code, d%)
- * f11a8c7 (update view, d%)
- * a21fdff (update html, d%)
- * cdb0571 (add html for buttons, d%)
- * 2744f71 (add number of students, d%)
- * a7af82b (add view, d%)
- * 624e6a3 (update inDegree, d%)
- * af55737 (Add inDegree, d%)

- * 670caf9 (add variable to stickfigure, d%)
- * 2df3021 (adjust csvStringToDat, d%)
- * 5a74bdc (update type Msg, d%)
- * d2fb4c5 (update main, d%)
- * d310ee9 (stickfigure construction, d%)
- * 3f6c387 (Add stickfigure, d%)
- * 9e35d79 (ap data with the defined variables, d%)
- * 81a3ee2 (plot points and description, d%)
- * 90529ff (add plot y axis, d%)
- * 4319f49 (plot x axis, d%)
- * 8b055ab (add svg for Stickfigure, d%)
- * be971c2 (add labelPositions, d%)
- * 9b9718b (Add parameters degree, d%)
- * 0fc818a (Add x snd yScaleLocal, d%)
- * 916b2d4 (declare functions and parameters, d%)
- * 41e0fcb (Add stickfigureplot, d%)
- * 7436557 (Add wideExtent, d%)
- * 5445d2c (Add defaultExtent, d%)
- * 285b0ba (Add yScale, d%)
- * 8f0284e (Add xscale, d%)
- * fd3551c (Add xAxis and yAxis, d%)
- * 98c707f (Add pointLabel, d%)
- * a973a81 (Add map data, d%)
- * 6078acd (Add filter function, d%)
- * b5627d8 (Add ChangeLen, d%)
- * b2aa113 (Add update, d%)
- * 06d72fa (map Data, d%)
- * 059dac2 (load Data, d%)
- * c6ff562 (Add type alias for StudentAcoholConsumption 13 values, d%)
- * 49f23b1 (Add main and init, d%)
- * ce89f5a (update VisualisationStickfigurePlot, d%)
- * a251a4f (Add VisualisationStickfigurePlot.elm, d%)
- * 82eb18d (update ParalleleKoordinaten, d%)

- * 3bdf7ee (Update d.dat, d%)
- * 1169ddf (update variable, d%)
- * 6ff365b (customize code, d%)
- * 4e1ec02 (install elm package, d%)
- * f86e1d7 (update Code, d%)
- * 2d347ce (Add svg, d%)
- * 9fd50b3 (Add decoding, d%)
- * 361942f (Add type MSG, d%)
- * c2f200b (Add Dataset, d%)
- * 708834d (update Code, d%)
- * aaffcdf (MultiDimDATA and MultiDimPOINT, d%)
- * f46e80a (type alias StudentAcoholConsumption, d%)
- * 47aa17c (type Model ParalleleKoordinaten, d%)
- * 25d7712 (update Packages ParalleleKoordinaten, d%)
- * d7f1e63 (update, d%)
- * fa729b7 (update colours, d%)
- * 25e8834 (Merge branch 'main' of https://github.com/Ella199/Elm_Project-Student-Alcohol-Consumption, d%)
- | * dfb2bcb (change variable "dalc" und "walc" in Dataset, d%)
- * | c1c6f1d (update Dataset ScatterplotTEST, d%)
- * | 7733b94 (Update colors, d%)
- * | d7dbf8e (update URL Scatterplot, d%)
- * | 6700404 (Merge branch 'main' of https://github.com/Ella199/Elm_Project-Student-Alcohol-Consumption, d%)
- | * 65854b8 (updated Dataset, d%)
- * | 779a950 (new Datasetupdate, d%)
- * | 3968b52 (Update Variable, d%)
- * | 66fe33d (Update, d%)
- * | 6a8bfdd (Update Model, d%)
- * | ffbae1c (first Test, d%)
- * | 680efdb (First Test Scatterplot for read Dataset, d%)
- * | 6ad75aa (update, d%)

- * 3fb07d7 (Merge branch 'main' of https://github.com/Ella199/Elm_Project-Student-Alcohol-Consumption, d%)
- | * 4b0cefd (changes in CSV, d%)
- * | f2bfead (update Dataset .Final NaN, d%)
- * | 46ea209 (update string, d%)
- * | 3b4727f (update, d%)
- * 8bdca1c (Dataset update, d%)
- * 8a3995a (code update, d%)
- * 7c7405e (changes update, d%)
- * c7026cf (read dataset update, d%)
- * 737e20c (update dataset Elm, d%)
- * 074d377 (Aufbereitung der Daten mit Python, d%)
- * 97fe74c (add Main, d%)
- * 6d9c59c (add new link, d%)
- * d133357 (abc, d%)
- * 437dc51 (change mergedstudent_Final to .csv, d%)
- * 204e23e (add mergedstudent_Final for Scatterplot, d%)
- * 49d1548 (Module importieren Scatterplot, d%)
- * 223912c (processed data sets, d%)
- * 30b9d1e (Initial commit, d%)arselas-MacBook-Air:Elm_Project-Student-Alcohol-Consumption
arselaleskaj\$
- * 624e6a3 (update inDegree, d%)
- * af55737 (Add inDegree, d%)
- * 670caf9 (add variable to stickfigure, d%)
- * 2df3021 (adjust csvStringToDat, d%)
- * 5a74bdc (update type Msg, d%)
- * d2fb4c5 (update main, d%)
- * d310ee9 (stickfigure construction, d%)
- * 3f6c387 (Add stickfigure, d%)
- * 9e35d79 (ap data with the defined variables, d%)
- * 81a3ee2 (plot points and description, d%)
- * 90529ff (add plot y axis, d%)
- * 4319f49 (plot x axis, d%)

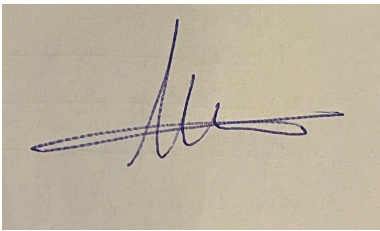
- * 8b055ab (add svg for Stickfigure, d%)
- * be971c2 (add labelPositions, d%)
- * 9b9718b (Add parameters degree, d%)
- * 0fc818a (Add x and yScaleLocal, d%)
- * 916b2d4 (declare functions and parameters, d%)
- * 41e0fcb (Add stickfigureplot, d%)
- * 7436557 (Add wideExtent, d%)
- * 5445d2c (Add defaultExtent, d%)
- * 285b0ba (Add yScale, d%)
- * 8f0284e (Add xscale, d%)
- * fd3551c (Add xAxis and yAxis, d%)
- * 98c707f (Add pointLabel, d%)
- * a973a81 (Add map data, d%)
- * 6078acd (Add filter function, d%)
- * b5627d8 (Add ChangeLen, d%)
- * b2aa113 (Add update, d%)
- * 06d72fa (map Data, d%)
- * 059dac2 (load Data, d%)
- * c6ff562 (Add type alias for StudentAcoholConsumption 13 values, d%)
- * 49f23b1 (Add main and init, d%)
- * ce89f5a (update VisualisationStickfigurePlot, d%)
- * a251a4f (Add VisualisationStickfigurePlot.elm, d%)
- * 82eb18d (update ParalleleKoordinaten, d%)
- * 3bdf7ee (Update d.dat, d%)
- * 1169ddf (update variable, d%)
- * 6ff365b (customize code, d%)
- * 4e1ec02 (install elm package, d%)
- * f86e1d7 (update Code, d%)
- * 2d347ce (Add svg, d%)
- * 9fd50b3 (Add decoding, d%)
- * 361942f (Add type MSG, d%)
- * c2f200b (Add Dataset, d%)
- * 708834d (update Code, d%)

- * aaffcdf (MultiDimDATA and MultiDimPOINT, d%)
- * f46e80a (type alias StudentAcoholConsumption, d%)
- * 47aa17c (type Model ParalleleKoordinaten, d%)
- * 25d7712 (update Packages ParalleleKoordinaten, d%)
- * d7f1e63 (update, d%)
- * fa729b7 (update colours, d%)
- * 25e8834 (Merge branch 'main' of https://github.com/Ella199/Elm_Project-Student-Alcohol-Consumption, d%)
- | * dfb2bcb (change variable "dalc" und "walc" in Dataset, d%)
- * | c1c6f1d (update Dataset ScatterplotTEST, d%)
- * | 7733b94 (Update colors, d%)
- * | d7dbf8e (update URL Scatterplot, d%)
- * | 6700404 (Merge branch 'main' of https://github.com/Ella199/Elm_Project-Student-Alcohol-Consumption, d%)
- | * 65854b8 (updated Dataset, d%)
- * | 779a950 (new Datasetupdate, d%)
- * | 3968b52 (Update Variable, d%)
- * | 66fe33d (Update, d%)
- * | 6a8bfdd (Update Model, d%)
- * | ffbae1c (first Test, d%)
- * | 680efdb (First Test Scatterplot for read Dataset, d%)
- * | 6ad75aa (update, d%)
- * 3fb07d7 (Merge branch 'main' of https://github.com/Ella199/Elm_Project-Student-Alcohol-Consumption, d%)
- | * 4b0cefd (changes in CSV, d%)
- * | f2bfead (update Dataset .Final NaN, d%)
- * | 46ea209 (update string, d%)
- * | 3b4727f (update, d%)
- * 8bdca1c (Dataset update, d%)
- * 8a3995a (code update, d%)
- * 7c7405e (changes update, d%)
- * c7026cf (read dataset update, d%)
- * 737e20c (update dataset Elm, d%)

- * 074d377 (Aufbereitung der Daten mit Python, d%)
- * 97fe74c (add Main, d%)
- * 6d9c59c (add new link, d%)
- * d133357 (abc, d%)
- * 437dc51 (change mergedstudent_Final to .csv, d%)
- * 204e23e (add mergedstudent_Final for Scatterplot, d%)
- * 49d1548 (Module importieren Scatterplot, d%)
- * 223912c (processed data sets, d%)
- * 30b9d1e (Initial commit,

Eigenständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen sind, sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.



Arsela Leskaj,

Wirtschaftsinformatik

Matrikelnummer: 216104691

Leipzig, 20.12.2022