

# Projektbericht zum Modul Information Retrieval und Visualisierung Sommersemester 2021

## Titel des Dokuments

Marcus Gagelmann

8. September 2021

### 1 Einleitung

Afrika bildet mit einer Größe von mehr als 30 Millionen Quadratkilometern den zweitgrößten Kontinent der Erde. Insgesamt leben dort über 1,3 Milliarden Menschen, was etwa 17,2% der Weltbevölkerung ausmacht. Wo jedoch Land und Menschen aufeinander treffen, dort sind auch Konflikte nicht weit entfernt. Seit vielen Jahren fällt Afrika solchen sowohl politischen als auch kriegesischen Auseinandersetzungen zum Opfer. Ursachen hierfür könnten die verbreitete Armut, misswirtschaftende Regierungen sowie die wertvollen Ressourcen des Kontinents sein, um nur einige mögliche Gründe zu nennen. Seit 1997 werden diese Konflikte durch das ACLED-Projekt dokumentiert. ACLED steht hierbei für „Armed Conflict Location and Event Data“. Die gesammelten Daten werden von ACLED auf ihrer Website (<https://acleddata.com/#/dashboard>) frei zugänglich zur Verfügung gestellt.

Nach nun beinahe 25 Jahren an Konfliktaufzeichnungen sind mittlerweile sehr große Datenmengen entstanden, welche einem außenstehenden Tabellenbetrachter kaum einen Überblick über das gesamtgeschehen seit dem Aufzeichnungsbeginn geben können. Noch unwahrscheinlicher ist es, dass ein solcher Betrachter allein mithilfe der unaufbereiteten Daten Rückschlüsse auf mögliche Zusammenhänge zwischen den vielen Konflikten ziehen kann. Mit dieser Ausgangslage ist ein Verstehen der tatsächlichen Lage in Afrika anhand der Daten unmöglich. Hierdurch könnten mögliche Maßnahmen zur Verbesserung der Situation des Kontinents weniger effektiv ausfallen, als wenn das volle Potenzial der Daten ausgeschöpft werden würde.

Aus diesem Problem ergibt sich die Fragestellung, ob der gewählte Datensatz hilfreiche Aussagen zu Zusammenhängen ermöglicht, welche die Opfer der Konflikte der letzten 25 Jahre

betreffen. Das Ziel dieser Arbeit ist deshalb die Bereitstellung einer Anwendung, welche eine Analyse der Todesopferanzahlen der Afrikakonflikte ermöglicht.

## **1.1 Anwendungshintergrund**

Mithilfe der im Rahmen dieser Arbeit entwickelten Anwendung werden einzelne Konflikte für einen Gesamtüberblick über eine oder mehrere Regionen Afrikas zusammengefasst, sowie nach einzelnen Jahren sortiert und hierbei auch mehrdimensional visualisiert. Insgesamt umfasst das Programm zwei verschiedene Ansichten, welche jeweils eine Menge von Konflikten auf unterschiedliche Art und Weise darstellen. In der Hauptansicht, dem Scatterplot, ist es dem Nutzer möglich einen Überblick über die Konfliktdichte und die Opferzahlen zu unterschiedlichen Zeitpunkten seit Aufzeichnungsbeginn zu erhalten. Die Zweitansicht, welche sich der Technik paralleler Koordinaten bedient, zeigt immer eine Teilmenge der Konflikte der Hauptansicht in jeweils drei Dimensionen. Hier kann der Nutzer einen genaueren Einblick erhalten, in welchem Monat eines Jahres welche Art von Konflikt wie viele Todesopfer forderte. Die Basismenge an Konflikten, welche für die Visualisierungen verwendet wird, ergibt sich aus der Menge aller Konflikte des Datensatzes, welche mithilfe eines Regionenfilters gekürzt wird. Dieser Filter wird vom Nutzer über eine dritte Ansicht in Form eines interaktiven Baumes gesteuert. Hier können sowohl die Konflikte größerer Regionen als auch die Konflikte einzelner Länder Afrikas in die visualisierte Gesamtmenge aufgenommen werden.

## **1.2 Zielgruppen**

Als Zielgruppe für die Visualisierungsanwendung kommen Personen in Frage, welche einen Überblick über die Konflikt- und Todesopferdichte zu verschiedenen Zeitpunkten und in verschiedenen Regionen des afrikanischen Kontinents erlangen wollen, sowie einen genaueren Einblick in die Art der Konflikte benötigen. Das Programm ist für einen schnellen Einstieg und einen groben Überblick in der Thematik der Afrikakonflikte konzipiert. Aus diesem Grund benötigen Zielgruppen in der Regel kaum Vorwissen um die Anwendung sinnvoll zu nutzen. Politisches und geografisches Vorwissen kann in manchen Anwendungsgebieten für das weitere Verständnis sinnvoll sein, ist aber keines Falls Voraussetzung um Erkenntnisse aus den Daten mithilfe der Anwendung zu gewinnen.

Die Verwendung einer solchen Anwendung ist bei Hilfsorganisationen zur Planung zukünftiger Einsätze denkbar. Hiermit könnten zukünftig diejenigen Regionen Afrikas ermittelt werden, welche langfristig die größten Opfer zu beklagen haben und aufbauend auf diesen Fakt weitere Hilfsleistungen am dringenden benötigen. Weiterhin können Hilfskräfte leicht einen Eindruck von der Art der bisherigen Konflikte einer bestimmten Region bekommen und damit zukünftige Einsätze besser auf diese Art von Konflikten vorbereiten.

Ein weiteres Anwendungsgebiet könnte die Risikoeinschätzung für Reisen durch das Auswärtige Amt (Link) darstellen. Hier bietet die Anwendung die Möglichkeit, die Gefahr für deutsche Reisende in Konfliktbelasteten Regionen Afrikas besser beurteilen zu können. Gäbe es in den letzten Jahren eine hohe Dichte an politischen Unruhen oder Todesopfer durch Kampfhandlungen, so können hierdurch Urlaubs- und Durchreisende rechtzeitig vor Gefahren gewarnt werden.

### 1.3 Überblick und Beiträge

Dieses Projekt setzt sich zusammen aus dem Datensatz des ACLED-Projekts von 1997-2021 und drei unterschiedlichen Visualisierungstechniken, welche interaktiv miteinander verbunden wurden. Vom verwendete Datensatz werden insgesamt sechs Felder jedes Datenobjektes für die Umsetzung der Anwendung genutzt. Hierzu gehören das Jahr sowie das Datum des Konflikts, die Art des Konflikts, die Region sowie das Land in dem ein Konflikt vorgefallen ist und die Anzahl an Todesopfern, welche ein Konflikt forderte. Die in dem Programm verwendeten Visualisierungstechniken beschränken sich auf eine Hauptansicht mit einem Scatterplot und einer umschaltbaren Zweitansicht, welche parallele Koordinaten zur Umsetzung verwendet. Zusätzlich gibt es noch einen Regionsfilter, welcher mithilfe expliziter Bäume anschaulich umgesetzt wurde. Im Scatterplot findet weiterhin die sogenannte X-Ray-Technik Anwendung, welche dem Nutzer Überlagerungen von Punkten aufzeigen kann.

Die Beiträge dieses Projekts belaufen sich auf die Auswahl, die Verarbeitung sowie die Visualisierung von Daten der Konflikte in Afrika in den Jahren 1997 bis 2021. Ganz konkret wird hierbei der Prototyp einer möglichen Anwendung diskutiert, welcher dem Nutzer die gewählten Daten und deren Zusammenhänge als interaktiv verbundenes System näher bringen soll. Hierbei werden die einzelnen Teile der Anwendung kritisch betrachtet, gelungene Ansätze hervorgehoben und Vorschläge zur weiteren Verbesserung geäußert.

## 2 Daten

Der Datensatz des ACLED-Projekt für die Afrikakonflikte von 1997 bis 2021 beinhaltet insgesamt 65443 einzelne Datenobjekte bzw. Datenzeilen, welche jeweils einen gesonderten Konflikt repräsentieren. Jedes Datenobjekt des Datensatzes liegt in insgesamt 29 Dimensionen vor, welche im einzelnen in Abbildung 1 eingesehen werden können. Verwendet werden von diesen Dimensionen in der Anwendung `EVENT_DATE`, `YEAR`, `EVENT_TYPE`, `REGION`, `COUNTRY` und `FATALITIES`. Diese gewählten Dimensionen beinhalten alle für das Programm benötigten Daten und bieten den Vorteil, unkodiert und fast ausschließlich in direkt verwendbarer Form vorzuliegen. Direkt verwendbar bedeutet hierbei, dass die Dimensionen keine weitere Verarbeitung benötigen und direkt in einfach Elm-Datenstrukturen (Int und String) umgewandelt werden können. Lediglich die Dimension `EVENT_DATE` bedarf einer zusätzlichen Übersetzerfunktion, welche genutzt werden kann, um den erhaltenen Event-Date-String auf den zugehörigen Monat

1. ISO	11. INTER1	21. LOCATION
2. EVENT_ID_CNTY	12. ACTOR2	22. LATITUDE
3. EVENT_ID_NO_CNTY	13. ASSOC_ACTOR_2	23. LONGITUDE
4. EVENT_DATE	14. INTER2	24. GEO_PRECISION
5. YEAR	15. INTERACTION	25. SOURCE
6. TIME_PRECISION	16. REGION	26. SOURCE_SCALE
7. EVENT_TYPE	17. COUNTRY	27. NOTES
8. SUB_EVENT_TYPE	18. ADMIN1	28. FATALITIES
9. ACTOR1	19. ADMIN2	29. TIMESTAMP
10. ASSOC_ACTOR_1	20. ADMIN3	

Abbildung 1: Dimensionen des ACLED-Datensatzes  
Quelle: Eigene Darstellung

zu mappen.

Es bleibt die Frage, in wie weit sich der Datensatz des ACLED-Projekts für die Zielgruppen dieser Arbeit eignet. Die eindeutigen Daten zu Zeitpunkt, Name des Landes und Todesopfern eignen sich gut für Hilfsorganisationen, um zu ermitteln, welche Länder aktuell am meisten unter den Konflikten leiden. Hierdurch kann eine Vorauswahl getroffen werden, welche Länder momentan Hilfsleistungen benötigen könnten. Mithilfe der Art des Events (EVENT\_TYPE) kann es diesen Organisationen weiterhin gut möglich gemacht werden, die Problemlage im Land besser zu verstehen und einen potenziellen Hilfseinsatz präziser auf eine bestimmte Art von Konflikten vorzubereiten. Bei der Aufgabe einer Gefahreinschätzung für Reisende durch das Auswärtige Amt kann mithilfe der gewählten Dimensionen des Datensatzes durchaus auch geholfen werden. Informationen wie die Konfliktdichte, die Todesopferdichte und auch die Konfliktarten in einer bestimmten Region lassen eine erste Einschätzungen zu der Situation eines Landes und zur Gefahr für Reisende zu.

Eine Verwendung der vorhandenen Geodaten zur Erweiterung der Anwendung durch eine interaktive Karte wäre durchaus denkbar und würde den Zielgruppen beim Verstehen von Zusammenhängen zwischen den Geodaten und den Konfliktdaten zusätzlich helfen. Eine solche Anwendung der Geodaten würde jedoch aufgrund von Problemen bei der Datenaufbereitung (siehe ??) sowie der nötigen Ergänzung des Projektes durch Polygondaten für Länderumrisse den Rahmen dieser Arbeit sprengen.

Es ist an dieser Stelle wichtig anzumerken, dass die Daten, welche im Programm Anwendung finden, nicht allein ausreichen um die genannten Aufgaben der Zielgruppen zu erfüllen. Da sowohl die Planung von Hilfseinsätzen, als auch die Einschätzung von Gefahrenländern Auswirkungen auf die Sicherheit von Menschenleben haben kann, ist eine Ergänzung des Datensatzes durch zusätzliche detaillierte Informationen zur Politik, Mentalität, Religion und anderen landes- und regionsspezifischen Informationen abseits der aufgezeichneten Konflikte unabdingbar. Beispielsweise kann offen gezeigte Homosexualität momentan in der Republik Kongo zu willkürlichen Verhaftungen wegen angeblich sittenwidrigen Verhaltens führen ([https://www.auswaertiges-amt.de/de/aussenpolitik/laender/kongorepublik-node/kongorepubliksicherheit/208542#content\\_1](https://www.auswaertiges-amt.de/de/aussenpolitik/laender/kongorepublik-node/kongorepubliksicherheit/208542#content_1)), was unbedingt bei einer Gefahreneinschätzung des Landes für westliche Reisende berücksichtigt werden sollte. Diese Information könnte jedoch nicht aus dem gewählten Datensatz ermittelt werden, da sich dieser lediglich auf größere Konflikte und nicht auf einzelne Verhaftungen innerhalb der Republik Kongo stützt.

## 2.1 Technische Bereitstellung der Daten

Bereitgestellt werden die verwendeten Daten von Kaggle ([urlhttps://www.kaggle.com/](https://www.kaggle.com/)), einer Plattform auf der unter anderem Datensätze für Data-Science-Projekte geteilt werden können. Eine andere Möglichkeit für den Download der Daten bietet die Seite des ACLED-Projekts ([urlhttps://acleddata.com/data-export-tool/](https://acleddata.com/data-export-tool/)). In dieser Arbeit wird der Datensatz von Kaggle verwendet, da die dort bereitgestellten Daten, anders als auf der ACLED-Seite, bereits nach Konflikten im Raum Afrika gefiltert sind. Lädt man den Datensatz direkt von Kaggle herunter, so erhält man diesen im CSV-Dateiformat. Als Separator werden hierbei Semikolons verwendet. Die Größe des Datensatzes beläuft sich auf etwa 30,5 MB (30.565.061 Bytes) im Standard CSV-Format, welche sich jedoch auf 62,4 MB (65.467.818 Bytes) mehr als verdoppelt nach einer Umwandlung der Daten in ein JSON-Dateiformat. Die Notwendigkeit für eine solche Umwandlung des Formats wird im Kapitel 2.2 genauer besprochen.

Jede Dimension des Datensatzes ist grundsätzlich entweder als Integer-Wert (Ziffern ohne Komma oder Anführungszeichen) oder als String-Wert (Zeichenfolgen in Anführungszeichen) gegeben. Hierzu gehören auch die Werte LATITUDE und LONGITUDE, welche eigentlich Float-Werte darstellen sollten. Dies stellt ein Problem, welches vor der eigentlichen Verwendung der Daten beim Einlesen durch eine Aufbereitung behoben werden müsste. Die Angaben in LATITUDE und LONGITUDE liegen als String vor, welcher unterschiedlich lange Ziffernfolgen enthalten kann. Hierbei ist keine Kommastelle angegeben, wodurch die Koordinaten des Konflikts nicht eindeutig wiedergegeben werden. Beispielsweise existiert ein Konflikt mit dem Breitengrad 36672 und dem Längengrad 2789. Mit diesen Daten könnten die Koordinaten (3.6672, 27.89) gemeint sein, welche eine Position im Wald in der Demokratischen Republik Kongo bestimmen. Aufgrund der Größe des Kontinents könnten hier jedoch auch die Koordinaten (36.672, 2.789) in Frage kommen, welche eine Position in der Stadt Koléa in Algerien beschreiben. In diesem Fall ist die zweite Möglichkeit korrekt. Es wird jedoch ersichtlich, dass die Position des Kommas in

den Breiten- und Längengraden nicht trivial bestimmt werden kann. Einige weitere Dimensionen des Datensatzes sind ebenfalls nicht sehr leicht zugänglich. Die Dimension INTERACTION liegt beispielsweise, neben einigen anderen, nur als numerischer Code vor. Um die tatsächliche Art der Interaktion der betroffenen Parteien ermitteln zu können, müsste hierfür vorher ein Übersetzer mithilfe des Codebooks des ACLED-Projekts ([https://reliefweb.int/sites/reliefweb.int/files/resources/ACLED\\_Codebook\\_2017FINAL%20%281%29.pdf](https://reliefweb.int/sites/reliefweb.int/files/resources/ACLED_Codebook_2017FINAL%20%281%29.pdf)) gebaut werden. Das Feld EVENT\_DATE, welches tatsächlich im Programm verwendet wird, besitzt ebenfalls eine erwähnenswerte Formatierung. Die Datumsangaben liegen nämlich im Format „TT-MMMMM-JJJJ“ vor, wobei der Monat in französischer Sprache ausgeschrieben wird. Hier werden an späterer Stelle (siehe Datenvorverarbeitung) weitere Datenverarbeitung nötig sein. Zuletzt bietet der gewählte Datensatz noch Metadaten, welche sich in die ungekennzeichnet in die anderen Datenfelder einordnen. Hierzu gehören die Dimensionen wie TIME\_PRECISION, welche einen Code für die Genauigkeit der Zeitangaben dieses Datenobjektes angibt und GEO\_PRECISION, welche einen Code für die Genauigkeit der Positionsangaben dieses Datenobjektes angibt. Auch diese müssten bei Bedarf mithilfe eines Übersetzers ausgelesen werden und anschließend zur Umwandlung einer Position/eines Zeitpunktes in einen Bereich/einen Zeitraum verwendet werden.

## 2.2 Datenvorverarbeitung

Im Rahmen dieses Projektes wurde die Entscheidung getroffen, den vorliegenden CSV-Datensatz in ein entsprechendes JSON-Dateiformat umzuwandeln. Zum einen bietet das JSON-Dateiformat den Vorteil für den Menschen besser lesbar zu sein, was sich als Hilfe bei manuellen Auswertungen während der Entwicklung des Programms herausstellte. Weiterhin werden bei der Konvertierung in das JSON-Format Objektstrukturen erstellt und im Datensatz gespeichert, welche von Elm beim Einlesen der Daten direkt in Records übersetzt werden können. Würde der Datensatz als CSV eingelesen werden, so müsste Elm zuerst die gesamte Datei als Zeichenkette einlesen und diese anschließend, aufgrund der fehlenden Objektstrukturen, aufwändiger während der Laufzeit parsen.

Beim Start des Programms müssen initiale Datenvorverarbeitungsschritte durchgeführt werden, um den View-Funktionen alle nötigen Daten zur korrekten Visualisierung bereustellen zu können. Hierzu gehört eine Initialfilterung der Regionen und Länder, um eine hilfreiche Startansicht gewährleisten zu können. Für diese Startansicht werden alle Konflikte herangezogen, welche sich seit 1997 im Land Ghana zugetragen haben. Hierfür wird die Funktion filterConflicts verwendet, welcher eine Liste aller Konflikte sowie alle aktiv geschalteten Länder und Regionen übergeben werden. Anschließend wird mithilfe der aktiven Länder und Regionen eine Liste derjenigen Länder erstellt, welche letztendlich angezeigt werden sollen. Dies ist wichtig, da in diesem Schritt auch die Länder ganzer aktiver Regionen zur ursprünglichen Liste aktiver Länder hinzu kommen können. Letztendlich wird mithilfe der resultierenden Liste aktiver Länder die Liste aller Kon-

flikte gefiltert. Hierbei werden alle Konflikte beibehalten, dessen Ländername mit einem Land aus der Liste aktiver Länder übereinstimmt.

Zusätzlich zur initialen geografischen Filterung müssen auch die französischen Datumsangaben der Dimension `EVENT_DATE` mit Tag, Monat und Jahr in die benötigten englische Monatsangaben übersetzt werden. Eine Übersetzung in ein auf den Tag genaues Datum wäre an dieser Stelle ebenfalls denkbar. Monatsangaben reichen hierbei jedoch von der Präzision her aus, um den Überblick innerhalb der parallelen Koordinaten (siehe später!!!) zu bewahren. Da die Dimension `EVENT_DATE` als String gegeben wird und die französischen Monatsnamen nicht variieren reicht hier eine einfache Übersetzerfunktion aus, welche einen `EVENT_DATE`-String auf die französischen Monatsnamen durchsucht und den zugehörigen englischen Monatsnamen zurück gibt. Mithilfe dieser Funktion kann anschließend eine Liste aller `EVENT_DATE`-Strings für die weitere Verwendung zu einer Liste mit englischen Monatsbezeichnungen gemappt werden.

Für die Visualisierung aller Konflikte eines einzelnen Jahres in der Zweitansicht als parallele Koordinaten wurde unter anderem die `EVENT_TYPE`-Dimension in der Anwendung dieses Projektes herangezogen. Hier wurde jedoch bewusst darauf verzichtet die untergeordnete `SUB_EVENT_TYPE`-Dimension ebenfalls zu visualisieren. Durch die Verwendung dieser Dimension würde die Zweitansicht des Programms zwar theoretisch einen höheren Detailgrad an Informationen zu den konkreten Geschehnissen eines Konflikts vermitteln. Es existieren jedoch 23 verschiedene Ausprägungen der `SUB_EVENT_TYPE`-Dimension, wodurch ein Nutzer bei Verwendung dieser Dimension leicht den Überblick verlieren könnte. Da jede einzelne Ausprägung dieser Dimension als relativ lange Zeichenkette in die Visualisierung eingebracht werden müsste, könnten sogar Überlagerungen der Achsenbeschriftungen auftreten. Eine Reduktion des Darstellungsaufwands ist an dieser Stelle keine Möglichkeit, weil jede einzelne Ausprägung der `SUB_EVENT_TYPE`-Dimension eine recht komplexe Information vermittelt, welche nur schwer abgekürzt werden kann. Ein Beispiel hierfür ist ein `SUB_EVENT_TYPE` des `EVENT_TYPE` „Battles“, welcher mit „Non-state actor overtakes territory“ bezeichnet wird.

## 3 Visualisierungen

### 3.1 Analyse der Anwendungsaufgaben

Insgesamt werden im Rahmen dieser Arbeit zwei Anwendungsaufgaben betrachtet. Zuerst soll die Anwendung Hilfskräften, welche zukünftige Einsätze in Afrika planen, einen schnellen aber gut verständlichen Überblick über die betroffenen Länder aktueller Konflikte verschaffen. Zusätzlich sollte es Mitarbeitern von Behörden wie dem Auswärtigen Amt mithilfe des Programms möglich sein, die Gefahrenlage für Reisende der westlichen Welt einschätzen zu können.

Für die erste Aufgabe ist es wichtig, dass die Zielgruppe gezielt Regionen und Länder auf

die zugehörige Todesopferdichte untersuchen kann. Weiterhin sollten die Nutzer des Programms in der Lage sein, die am häufigsten vorkommenden Konfliktarten prüfen zu können. Diese Informationen sind nötig, um eingrenzen zu können, welche Regionen und Länder zusätzliche Hilfeleistungen am dringenden benötigen. Weiterhin ist es Nötig die durchschnittliche Anzahl an Todesopfern bei einem Konflikt eines bestimmten Landes oder einer Region einschätzen zu können, um die Quantität der zukünftigen Hilfseinsätze festzulegen. Zuletzt ist ein Überblick über die häufigsten Konfliktarten notwendig, damit Hilfsorganisationen einschätzen können, ob und wie viele Sicherheitskräfte einen eventuellen Hilfseinsatz begleiten müssen.

Die nötigen genannten Informationen können zum Teil aus der Anwendung gewonnen werden. Beispielsweise hilft der Scatterplot in Verbindung mit dem Regionsfilter beim Aufbau eines mentalen Modells zur Todesopferdichte der verschiedenen Regionen in den letzten Jahren. Hierbei verinnerlicht der Nutzer einen Eindruck von denjenigen Jahren, in denen sich besonders viele Konflikte überlagern, sowie einzelnen „ausreißenden“ Konflikten, welche eine besonders hohe Anzahl an Todesopfern forderten. Mithilfe von Parallelen Koordinaten können anschließend einzelne Jahre noch präziser betrachtet werden, um den ersten Überblick über die Opfer dieses Jahres zu festigen und grob einschätzen zu können, welche Art von Konflikten die meisten Opfer forderte. Kritisch betrachtet werden sollte die Fähigkeit der Anwendung, die durchschnittliche Todesopferanzahl eines Jahres zu vermitteln. Mithilfe der verwendeten X-Ray-Technik des Scatterplots sind zwar Konflikthäufungen auszumachen, eine genaue Einschätzung des Todesopferdurchschnitts ist jedoch durch die Verwendung absoluter Zahlen und Überlagerungen nur bedingt möglich. Hier sollte entweder die Anwendung mithilfe weiterer Ansichten ausgebaut, oder externe Quellen zur Ergänzung des Programms herangezogen werden. Nicht nötig sind für diese Anwendungsaufgabe die Jahre, welche sich bereits weit in der Vergangenheit befinden. Hier können keine Informationen abgelesen werden, welche aktuelle Hilfseinsätze beeinflussen würden.

Für die Aufgabe der Gefahreneinschätzung durch das Auswärtige Amt, welche die Anwendung unterstützen soll, wird ein ähnliches mentales Modell wie für die erste Anwendungsaufgabe benötigt. Hier ist ein breiter Überblick über die Konfliktdichte eines Landes in den letzten Jahren wichtig, um effizient Kandidaten unter den Ländern auszumachen, welche anschließend gründlich auf eine potenzielle Einreisewarnung untersucht werden müssen. Für diese Anwendungsaufgabe könnten weit vergangene Jahre ebenfalls interessant sein, um die Konfliktgeschichte eines Landes besser nachempfinden zu können und damit das zukünftige Gefahrenpotenzial besser einschätzen zu können.

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

Im Rahmen dieser Arbeit soll eine Anwendung entstehen, welche eine unterstützende Rolle bei Analysen der Afrikakonflikte von 1997-2021 einnimmt. Die Hauptanforderung an das Programm ist hierfür, dem Nutzer einen schnellen aber effektiven Überblick darüber zu geben, welche Re-



gionen Afrikas besonders von Konflikten geprägt sind und welche Todeszahlen sich hieraus im Vergleich zu anderen Regionen ergeben. Zu diesem Zweck sollten die unterschiedlichen Visualisierungsansichten des Programms in einem relativ „leichten“ Design gehalten werden. Eine Überzahl an gleichzeitig gezeigten Informationen könnte zwar den Detailgrad der Anwendung erhöhen, der eigentliche Sinn der Anwendung würde hierdurch jedoch in den Hintergrund gerückt werden.

Die Filterfunktion für Regionen und Länder ist einer der wichtigsten Teile der Anwendung und bedarf besonderer Sorgfalt bei der Entwicklung. Dem Nutzer muss es möglich sein, flexibel einzelne Länder und ganze Regionen Afrikas zum Filter für die anderen Visualisierungen hinzuzufügen. Zusätzlich sollten Nutzer der Anwendung jeder Zeit Feedback dazu bekommen, welche Konflikte aus welchen Ländern und Regionen ihm momentan angezeigt werden. Zuletzt ist es für den Nutzer durchaus sinnvoll, permanent die geografischen Zusammenhänge zwischen den filterbaren Ländern und Regionen vermittelt zu bekommen, um einen besseren Überblick über aktuelle Filtereinstellungen und deren Bedeutung zu bewahren (siehe später !!!).

Die Hauptaufgabe der Anwendung ist die Vermittlung eines Eindrucks der schwere von Konflikten in verschiedenen Ländern und Regionen. Zu diesem Zweck sollte unbedingt eine Darstellung der Opferzahlen in Relation zu unterschiedlichen sinnvollen Dimensionsausprägungen erfolgen, um einzelne Ereignisse, ganze Länder und Länderübergreifende Regionen zu einem bestimmten Zeitpunkt besser zu verstehen. Hierfür bieten sich beispielsweise Zeiteinheiten, Konfliktarten oder auch Geodaten an. Die Verwendung der Jahresdaten ist in diesem Fall besonders entscheidend für die Zielgruppen, um aktuelle Ereignisse von weit vergangenen unterscheiden zu können. Die genauere Untersuchung einzelner Jahre als „visualisierter Ausschnitt“ könnte hierbei von Vorteil sein.

### **3.3 Präsentation der Visualisierungen**

Präsentieren sie die visuelle Abbildungen und Kodierungen der Daten und Interaktionsmöglichkeiten. Sie müssen begründen, warum und wiegut ihre Designentscheidungen die erstellten Anforderungen erfüllen. Weiterhin müssen sie begründen, warum die gewählte visuelle Kodierung der Daten für das zulösende Problem passend ist. Typische Argumente würden hier auf Wahrnehmungsprinzipien und Theorie über Informationsvisualisierung verweisen. Die besten Begründungen diskutieren explizit die konkrete Auswahl der Visualisierungen im Kontext von mehreren verschiedenen Alternativen. Diskutieren sie die Expressivität und die Effektivität der einzelnen Visualisierungen.

Die eben beschriebenen Präsentationen und Begründungen sollen für jede der drei folgenden Visualisierungen durchgeführt werden.

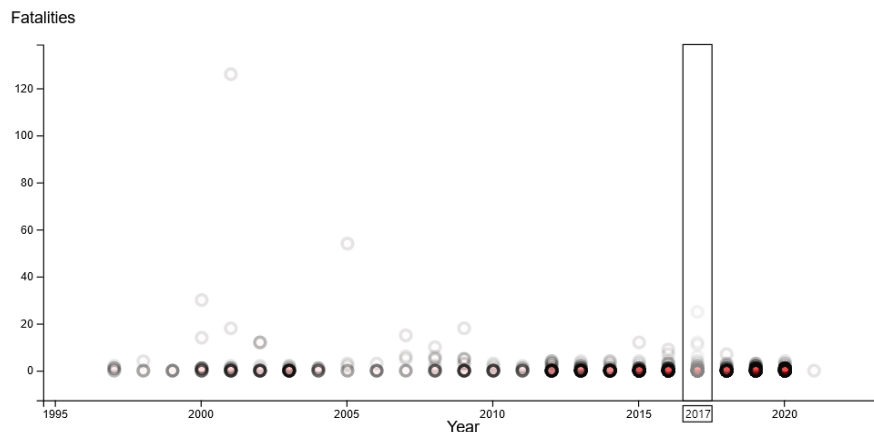


Abbildung 2: Screenshot eines möglichen Scatterplots  
Quelle: Eigene Darstellung

### 3.3.1 Visualisierung Eins

Die erste Visualisierung innerhalb der Anwendung ist beispielhaft in Abbildung 2 zu sehen und zeigt einen zweidimensionalen Scatterplot, welcher sowohl die Dimension der Todesopferanzahlen als auch die zugehörigen Jahre darstellt. Hierbei werden die Achsen dynamisch an die aktuell angezeigten Daten angepasst, um diese immer maximal auf die zur Verfügung stehende Fläche zu verteilen. Jeder Datenpunkt in der Visualisierung stellt einen Konflikt des ACLED-Datensatzes dar. Diese werden in der Darstellung mit einer hohen Transparenz angezeigt, um einen X-Ray-Effekt zu erzeugen, wodurch Überlappungen leicht erkannt werden können. Das Design der Datenpunkte ist geprägt durch einen dicken schwarzen Ring mit einer hohen Transparenz (alpha-Wert von 0.1), welcher mit einem Rotton gefüllt ist, der nochmal um ein zehnfaches transparenter ist (alpha-Wert von 0.01). Das letzte Detail der Scatterplot-Visualisierung ist das interaktive Feature der Jahresauswahl. Hierfür wird der Scatterplot in gedachte Spalten eingeteilt, welche an Jahresachse ausgerichtet sind. Durch einen Hover der Maus über eine dieser Spalten erscheint für die Dauer des Hovers ein Spaltenumfassendes Rechteck mit schwarzen Rändern und stark transparenter weißer Füllung. Zusätzlich erscheint auf der Jahresachse das betreffende, schwarz umrandete Jahr für die Dauer des Maus-Hovers. Ein Klick auf eine der markierten Flächen lässt die Anwendung in die zweite Visualisierung (siehe !!!) des ausgewählten Jahres wechseln.

Die Anforderung an die Visualisierung, einen ersten schnellen Überblick über die Todesopferzahlen der angezeigten Regionen zu gewähren, wird von der ersten Ansicht größtenteils erfüllt. Grund hierfür ist das minimalistische aber leicht verständliche Design, welches auf der allgemein bekannten Visualisierungstechnik des zweidimensionalen Scatterplots beruht. Dieser Fakt in Kombination mit nur zwei unterschiedlichen verwendeten Farben ermöglicht es einem Betrachter ohne längere Betrachtungszeit einen Überblick über die präsentierten Informationen zu

gewinnen. Der Gehalt der präsentierten Informationen könnte wiederum noch höher sein.

Ein Problem der ersten Visualisierung ist die Umsetzung der X-Ray-Technik. Diese vermittelt zwar eindeutig und auf den ersten Blick höhere und weniger hohe Konfliktichten, wie viele Datenpunkte in solchen Clustern jedoch genau vorhanden sind kann der Betrachter nur schlecht erahnen. Ab einer gewissen Konfliktüberlappung gleichen sich die Cluster zu stark um sie präzise unterscheiden zu können. Dies könnte seinen Ursprung darin haben, dass ein linearer Anstieg des Helligkeitsstimulus nur zu einem sublinearen Zuwachs an wahrgenommener Helligkeit (oder Dunkelheit) korrespondiert (zitat S2S130 ca). In anderen Worten bedeutet das, dass ab einer gewissen Sättigungs- bzw. „Dunkelheitsstufe“ bei einem Datenpunkte bei gleich bleibender, linearer verdunklung immer weniger Veränderung für das menschliche Auge wahrnehmbar ist. Hier könnte die Anwendung einer Gammakorrektur abhilfe verschaffen, welche die linearen Helligkeitsabstufungen auf eine andere gewünschte Weise skalieren kann. Hiermit könnten auch bei höheren Datenpunktdichten noch wahrnehmbare Unterschiede herbeigeführt werden.

Ein weiteres Problem des Scatterplots ist die Verteilung der Datenpunkte auf der Fatalities-Achse. Hier ist bei fast allen möglichen Regionszusammensetzungen zu beobachten, dass sich der Großteil der Datenpunkte im untersten Bereich der Grafik ansammeln. Einzelne Ausreißer lassen die Reichweite der Fatalities-Achse jedoch bis zu weit überdurchschnittlichen Werten verlaufen. Um die Datenpunkte auf dieser Grundlage besser in der Grafik zu verteilen, könnte hier eine logarithmische statt einer linearen Skalierung der Fatalities-Achse angebracht sein. Hierdurch würden die stark komprimierten Datenpunkte um unteren Ende der Grafik eventuell besser im oberen Bereich verteilt werden. Hierdurch würde zum einen der Platz innerhalb der Grafik besser zur Informationsdarstellung genutzt werden, zum anderen könnte dieser Ansatz ebenfalls beim ersten Problem der zu hohen Datenpunktdichte helfen.

### 3.3.2 Visualisierung Zwei

Die zweite Visualisierung des Projektes zeigt einige Dimensionen von Konflikten aus ausgewählten Regionen, dargestellt als parallele Koordinaten. Zu diesem Zweck wurden drei parallele Achsen angelegt, eine für jede Datendimension, wie in Abbildung 3 zu sehen ist. Auf die erste dieser Achsen wird die Dimension `EVENT_DATE` einzelner Konflikte abgebildet, welche in der Datenvorverarbeitung (siehe 2.2) bereits auf den Monat, in welchem das Ereignis stattgefunden hat, reduziert wurde. Die anderen beiden Achsen beschreiben, ohne weitere Datenaufbereitung, die Anzahl der Todesopfer bei einem Konflikt (`FATALITIES`) und die Art eines Konflikts (`EVENT_TYPE`). Einzelne Konflikte werden in dieser Visualisierung als dünne schwarze Pfade dargestellt. Diese verlaufen von ihrem zugehörigen Monat auf der linken Seite über ihr zugehörige Todesopferanzahl in der Mitte bis hin zu ihrer zugehörigen Konfliktart auf der rechten Seite. Das aktuell eingestellte Jahr für diese Visualisierung ist in der oberen linken Ecke zu erkennen. Aufgrund der höheren Anzahl an Datendimensionen, welche sich größtenteils von denen der ersten Visualisierung unterscheiden, besitzt diese Darstellung einen höheren und anderen Infor-

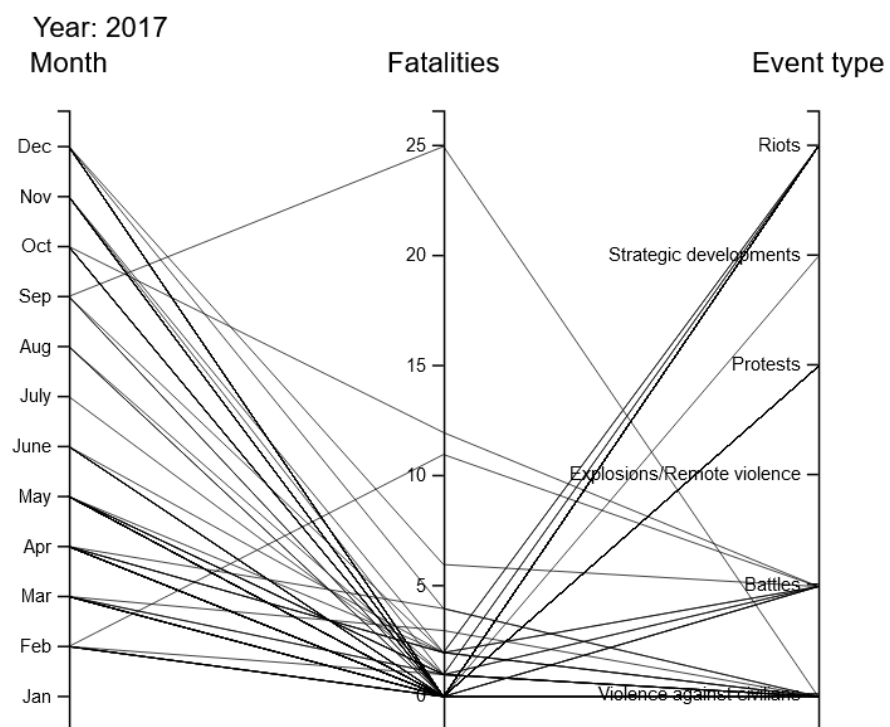


Abbildung 3: Screenshot einer möglichen Darstellung paralleler Koordinaten  
 Quelle: Eigene Darstellung

mationsgehalt im Vergleich zur ersten. Trotzdem ist die verwendete Anzahl an Dimensionen mit lediglich dreien immernoch relativ gering. Zum einen trägt dies zu dem Anwendungsziel bei, die Visualisierungen für das schnelle Verstehen durch den Nutzer zu designen. Eine insgesamt recht geringe Dimensionsvielfalt zusammen mit der Verwendung von nur einer Farbe und einer oft nur geringen Anzahl an visualisierten Konflikten sorgt in dieser Darstellung beim Nutzer für einen schnellen Überblick, um effizient mehrere Versionen der zweiten Visualisierung zu verstehen. Es ist jedoch zu beachten, dass aufgrund dieses Minimalismus Potenzial verloren geht, dem Nutzer Informationen in größerer Qualität und Quantität zu vermitteln. Mithilfe der Achse der Dimension `EVENT_DATE` hätte beispielsweise auch das gesamte Jahr in Tagesabstufungen und nicht in Monatsabstufungen abgebildet werden können. Außerdem gibt es durchaus weitere interessante Dimensionen (`SUB_EVENT_TYPE`, `ACTOR1`, `ACTOR2`, etc.), welche man zusammen mit den dargestellten Daten in Bezug setzen könnte. Die verwendeten Dimensionen der zweiten Visualisierung wurden ausgewählt, um ein besseres Verständnis beim Nutzer zu schaffen für den Zeitpunkt und die Umstände, welche zu der zugehörigen Anzahl an Todesopfern während eines Konfliktes führten. Die `FATALITIES`-Achse wurde in der Mitte platziert, um zu beiden anderen Dimensionen in direkter Relation zu stehen und eine räumliche Nähe zu den anderen Achsen zu schaffen.

Die zweite Visualisierung erfüllt die Aufgabe, dem Nutzer einen genaueren Ausschnitt aus der Menge aller Konflikte der gewählten Region zu beschreiben. Hierfür wird das sogenannte „Slicing“ verwendet, welches eine Methode der Datenanalyse darstellt. Dabei wird aus einem Datenwürfel eine Scheibe „herausgeschnitten“. In diesem Fall wird ganz konkret die Dimension `YEAR` der ersten Visualisierung auf ein einzelnes Jahr eingeschränkt, was gut in Abbildung 2 durch die zugehörige Hover-Visualisierung verdeutlicht wird.

Ein Nachteil dieser Visualisierung ist, ähnlich wie schon beim Scatterplot, dass die Untersuchung einzelner Konflikte hier nur in Ausnahmefällen möglich ist. Oft ist nicht genau erkennbar, wie viele Konflikte sich auf einer Linie überlagern. Bei Betrachtung der gesamten Abbildung wird aber durchaus deutlich, in welchen Ausprägungen der verschiedenen Dimensionen hohe oder niedrige Konfliktdichten vorliegen. In Abbildung 3 ist beispielsweise zu sehen, dass hier die meisten Konflikte ungefähr gleich auf die Monate verteilt liegen, die meisten Konflikte weniger als fünf Todesopfer forderten und kein einziger Konflikt der Art „Explosions/Remote violence“ war. Das Anwendungsziel des Gesamtüberblicks bleibt also durchaus erhalten. Um einen besseren Überblick über einzelne Konflikte zu behalten könnten in dieser Visualisierung verschiedene Farben für verschiedene Konfliktarten oder aber auch, wie schon oben genannt, präzisere Skalen helfen, welche weniger Konfliktpfade auf ein und derselben Linie entstehen lassen würden.

### 3.3.3 Visualisierung Drei

Die dritte und letzte Visualisierung liegt in der Form eines Baumes vor, wie in Abbildung 4 zu erkennen ist. Dieser Baum soll die verschiedenen, von Konflikten betroffenen Regionen Afrikas

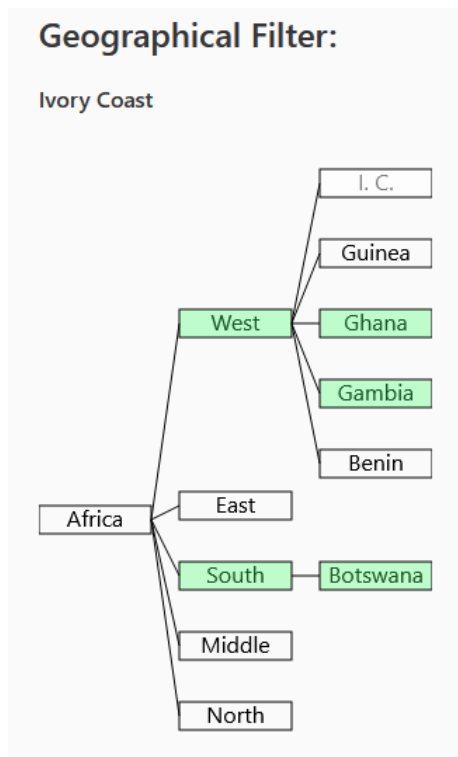


Abbildung 4: Screenshot eines möglichen Filterbaumes mit Maus-Hover über dem „I. C.“-Knoten  
Quelle: Eigene Darstellung

und deren geografisch-hierarchische Strukturen verdeutlichen. Die Knoten des Baumes bestehen aus Rechtecken mit schwarzem Rand und weißem Hintergrund, auf welchen der Name der zugehörigen Region zentral geschrieben steht. Verbunden sind diese Knoten durch scharze Linien, welche die Kanten des Baumes repräsentieren. Der Baum ist in drei Ebenen eingeteilt, welche es als Nutzer zu verstehen gilt. Diese sind von links nach rechts vertikal ausgerichtet und beginnen in der ersten Ebene mit lediglich der Wurzel, welche das Wort „Afrika“ enthält. Von diesem Afrikaknoten zweigen immer fünf Kanten nach rechts in die zweite Ebene ab und enden in Knoten, welche verschiedene Region Afrikas repräsentieren. Zuletzt kann auch noch jeder Regionsknoten in die dritte Ebene nach rechts ausgeklappt werden, um Knoten der verschiedenen Konfliktbetroffenen Länder der zugehörigen Region zu zeigen. Hierüber bietet die Visualisierung interaktive Funktionen zur Filterung der in den anderen Darstellungen präsentierten Konflikte nach Regionen und Ländern an. Aktiv geschaltene Knoten werden hierbei grün von der ansonsten schwarz-weißen Darstellung hervorgehoben. Genauer hierzu wird im Kapitel 3.4 beschrieben.

Auch diese Visualisierung ermöglicht dem Nutzer effizient an einen Überblick über die dargestellte Information zu gelangen, diesmal durch eine geringe Baumbreite und -tiefe, ein minimalistisches Komponentendesign mit wenigen unterschiedlichen Formen und Farben sowie ausklappbaren Teilbäumen, welche die Informationsflut auf den Nutzer so gering wie möglich halten.

Ursprünglich war für die Filterfunktion der Anwendung eine Liste von Checkboxes angedacht, welche sich jedoch als problematisch erwies. Bei 18 darzustellenden Ländern entstand eine zu lange Liste, um schnell gesuchte Länder finden zu können. Hierzu kommt die Tatsache, dass auf diese Weise zusätzlich keine gute Filterfunktion für gesamte Regionen existierte. Um dies zu ermöglichen hätte die Liste von Checkboxes unter Umständen noch länger werden müssen und noch weniger Überblick geboten. Es ist also zu erkennen, dass die Gewählte Möglichkeit einen hierarchischen Filter mithilfe eines Baumes zu realisieren und zu visualisieren durchaus Vorteile in diesem Anwendungsfall mit sich bringt.

Es besteht jedoch weiterhin auch ein Nachteil im Zusammenhang mit längeren Regions- und Ländernamen. Diese werden zum Teil zu lang um sie in einen Knoten einer Ebene zu schreiben. Für dieses Problem bieten sich verschiedene Lösungsansätze. Unter anderem wäre es denkbar die Knotenhöhe zu vergrößern und lange Namen zweizeilig zu schreiben. Eine Andere Möglichkeit wäre es, den Baum und die Knoten weiter in die Breite zu vergrößern und somit die Namen einzeilig ausschreiben zu können. Beide Möglichkeiten nehmen jedoch zu viel Platz der Gesamtanwendung in Anspruch, wodurch der Filterbaum auf eine eigene Seite oder zumindest in ein eigenes Modal verlegt werden müsste. Hierdurch wäre es dem Nutzer jedoch nicht mehr möglich zu jedem Zeitpunkt die aktuellen Filtereinstellungen sowie die Regions- und Länderhierarchie einsehen zu können, wodurch wichtige Informationen beim Verstehen der anderen Visualisierungen verloren gehen. Da dieser Ansatz offensichtlich gegen das Anwendungsziel der Übersichtlichkeit arbeitet, wurde das Problem auf anderem Weg umgangen. Die zu langen Namen in den Knoten des Baumes werden im Programm durch eine Elm-Funktion automatisch auf

ihre Anfangsbuchstaben reduziert (z.B. „I. C.“ für Ivory Coast). Beim Verstehen dieser Abkürzungen wird der Nutzer durch eine Maus-Hover-Funktion unterstützt, welche den vollen Namen eines Knotens oberhalb des gesamten Baumes anzeigt.

### 3.4 Interaktion

-scatterplot: spaltenauswahl -> ansichtwechsel + Bild -pc: buttons: zurück, jahr zurück, jahr vor + bild -tree: filter genauer erklären (wann wird welches feedback gegeben und welche konflikte werden dadurch angezeigt?) -checkboxes als filter nicht umgesetzt -> bessere übersicht/schnelleres finden über geografische hierarchie

## 4 Implementierung

-in elm -> -aufbau: module kurz beschreiben -> imports erklären -> css-file + bulma -wichtigste datenstrukturen (conflict, geotree, filter, Model) erklären: aufbau, funktion -implemtierungsaufwand: grundsätzliches übernehmen bei allen dreien aus übungsaufgaben -> aufwand lag darin code aufzuräumen und code an neue funktion anzupassen (z.B. visual auf neue datenstrukt zuschneiden) -> besonders aufwändig war anpassung der scalen von pc an ordinale daten

-1. impl: view=renderScatterplot-fkt -> bekommt liste conflicts -> führt achsen, punkte und yearSelectionBoxes (mehr siehe später dieses kapitel) zusammen

-2. impl: view=renderParallelCoordinates-fkt -> bekommt liste conflicts + year (int) -> zeichnet aus gesonderten fkt ordinale achsen (yAxisOrdinal) + kontinuierende achsen (yAxisFloat) -> filtert konflikte nach übergebenem jahr, zeichnet konflikt-pfade mithilfe einzelner linien von achse zu achse; interessant beim auslesen der relevanten werte der konflikte: übersetzerfunktion (dateStringToMonthString) (+codebild)

-3. impl: view= renderTree-fkt -> bekommt geotree + filter -> verwendet buildTree-fkt welche den eigentlichen Baum zusammensetzt: countrynodes + regionnodes erklären (codebild) cns: liste aller regionen mit zugehörigen ländern -> eine region mit zugehörigen ländern -> Tree.node für jedes Land als blatt rns: für jede region: durchsuche cns-dict nach passendem region-key -> erstelle node für diese region -> füge cn-liste des zugehörigen cn-dict-eintrages als kinder ein root: node mit afrika und rns als kinder filter des trees: filter an jeden drawnode-aufruf übergeben -> gucken ob land oder region des knotens in filter -> falls aktiv: css class activeNodeBox wodurch node eingeblendet wird

-interaktion: yearSelectionBoxes + codebild, bulma-buttons, custom tree buttons mit css

Beschreiben Sie die Implementierung ihrer Visualisierungsanwendung in Elm. Stellen die Gliederung ihres Quellcodes vor. Haben Sie verschiedene Elm-Module erstellt. Was war aufwändig umzusetzen, was ließ sich mit dem vorhanden Code aus den Übungen relativ einfach umsetzen?

Wie sieht die Elm-Datenstruktur für das Model aus, in dem die verschiedenen Zustände der Interaktion gespeichert werden können.



## **5 Anwendungsfälle**

Präsentieren sie für jede der drei Visualisierungen einen sinnvollen Anwendungsfall in dem ein bestimmter Fakt, ein Muster oder die Abwesenheit eines Musters visuell festgestellt wird. Begründen sie warum dieser Anwendungsfall wichtig für die Zielgruppe der Anwenderinnen ist. Diskutieren sie weiterhin, ob die oben beschriebene Information auch mit anderen Visualisierungstechniken hätte gefunden werden können. Falls dies möglich wäre, vergleichen sie die den Aufwand und die Schwierigkeiten ihres Ansatzes und der Alternativen.

### **5.1 Anwendung Visualisierung Eins**

### **5.2 Anwendung Visualisierung Zwei**

### **5.3 Anwendung Visualisierung Drei**

## **6 Verwandte Arbeiten**

Führen sie eine kurze Literatursuche in der wissenschaftlichen Literatur zu Informationsvisualisierung und Visual Analytics nach ähnlichen Anwendungen durch. Diskutieren sie mindestens zwei Artikel. Stellen sie Gemeinsamkeiten und Unterschiede dar.

## **7 Zusammenfassung und Ausblick**

-verschiedene Länder/Regionen unterscheiden (z.B. Farben)

Fassen sie die Beiträge ihre Visualisierungsanwendung zusammen. Wo bietet sie für die Personen der Zielgruppe einen echten Mehrwert.

Was wären mögliche sinnvolle Erweiterungen, entweder auf der Ebene der Visualisierungen und/oder auf der Datenebene?

## **Anhang: Git-Historie**