

Projektbericht zum Modul Information Retrieval und  
Visualisierung Sommersemester 2023

# **Visualisierungen zur Analyse von Einflussfaktoren auf die Schlaf Effizienz**

Mick Stewart Wörner

27. Dezember 2023

# 1 Einleitung

Tipps zu Latex und Koma-Script für Hausarbeiten sind im LaTeX Reference Sheet for a thesis with KOMA-Script von Marion Lammarsch und Elke Schubert zusammengefasst. Der Bericht fällt in die Kategorie von InfoVis-Paper, die Tamara Munzner Design Study nennt [Munzner2008]: In der Einleitung sollen sie zuerst das Zielproblem beschreiben. Daraus sollen sie Fragestellungen motivieren, die mittels Techniken der Informationsvisualisierung beantwortet werden können. In dem Abschnitt direkt unter der Überschrift Einleitung sollen Sie nach einer kurzen Einleitung Fragestellungen und das Zielproblem motivieren und beschreiben.

## 1.1 Anwendungshintergrund

Sie müssen genug Hintergrund bereitstellen, so dass die Lesenden sich ein Urteil bilden können, ob ihre Lösung funktioniert. Sie sollen die Lesenden jedoch nicht mit Anwendungsdetails so überschütten, dass der Fokus auf die Fragen zur Informationsvisualisierung untergehen. Eine Visualisierung muss **expressiv: Expressivität bedeutet, dass die Daten unverfälscht wiedergegeben werden. Grundsätzlich sollen nur die Informationen dargestellt werden, die auch im Datenmaterial vorhanden sind.**

I effektiv (Effektivität hängt nicht nur von den Daten ab, sondern auch von:

1. dem Bearbeitungsziel und
2. den Fähigkeiten des Betrachters
3. Eine effektive Visualisierung versucht die Inhalte auf intuitiven Wegen zu präsentieren) I angemessen sein (Angemessenheit beschreibt den Verbrauch an Ressourcen zur Erzeugung der Visualisierung)

## 1.2 Zielgruppen

Die Schlafqualität hat einen signifikanten positiven Einfluss auf die Lebensqualität von Menschen. [2] Das Menschliche schlafsystem ist allerdings ein hochkompliziertes System das von Multiplen externen Faktoren beeinflusst werden kann. Die in diesem Projekt dargestellten Visualisierungen sollen dabei helfen Einflüsse und Zusammenhänge bei der Schlafqualität zu erkennen und Forschern sowie Privatpersonen dabei zu helfen den Einfluss von Lebensstilen auf verschiedene Maße von Schlafqualität zu erkennen. Die untersuchten Attribute sind der Koffein, Alkohol, Tabak Konsum und Sport sowie das Alter und Geschlecht der Personen. Dies sollte Personen helfen die Verhaltensweisen zu identifizieren, mit denen Sie den größten Einfluss auf Ihre Schlafqualität haben könnten. Dabei kann ausgegangen werden, dass die Benutzer wissen, dass die Datenlage kritisch in Hinblick auf die Aussagekräftigkeit und praktische Implikationen reflektiert werden muss. Dies ist wichtig, da der benutzte Datensatz  $n < XXXX$  Datenpunkte besitzt, welche durch den eingebauten Filter weiter reduziert werden können. ab  $n < 30$  ist die Statistische Aussagekraft

**nichtmehr gegeben. klein ist.** Beschreiben sie die Personengruppe oder Personengruppen, die das von ihnen benannte Anwendungsproblem lösen möchte. Auf welches Vorwissen können sie in dieser Gruppen von Anwenderinnen aufbauen? Welche Informationsbedürfnisse werden durch die Visualisierungen adressiert?: **Überarbeiten:**

**The data was then analyzed to understand the relationship between lifestyle factors and sleep patterns and to identify any potential areas for intervention to improve sleep**

### 1.3 Überblick und Beiträge

**Im diesem Abschnitt wird eine Überblick auf die Daten und verwendeten Visualisierungstechniken.** In diesem Abschnitt geben sie einen kurzen Überblick über die Daten und verwendeten Visualisierungen. Dann benennen sie die Beiträge ihres Projekts. Diese Beiträge müssen sie in den hinteren Teilen des Berichts genauer ausführen und belegen.

## 2 Daten

Beschreiben Sie vorhandenen Daten. Der verwendete Datensatz besteht aus: 452 Personen, welche durch Ihre ID identifiziert werden. Ob die ID nur den Datenpunkt oder die Person identifiziert ist unklar. Daher lässt sich nicht sagen ob Schlafdaten einer Person mehrfach erfasst wurden sind. Auf der Kaggle Seite wurde nach Angaben des Authors erwähnt, dass der Datensatz im Kontext einer Studie von der ENSIAS, Marroco gesammelt wurde. Innerhalb einer eingeschränkten Recherche konnten weder auf der Webseite der ENSIAS noch in weitergehender Literaturrecherche eine Quelle identifiziert werden. Daher sollten die Daten und daraus entwickelten Ergebnisse, nicht unreflektiert übernommen werden. Der Datensatz hat 15 Attribute, Nominale (Id, Gender und Raucher) und Quantitative (Rest) .

Erste Gruppe: Identifikatoren, sind zur eindeutigen Bestimmung eines Datentupels oder Person. Dieser Gruppe gehört nur die "ID" an. Die ID identifiziert eine Person einmalig. Da keine ID wurde mehrfach aufgeführt wird ist anzunehmen, dass jede Person nur einmalig an der Studie teilgenommen hat. Die ID wird als Integer bereitgestellt. Der Datenbereich geht von [1-452]

Age: Gibt das Alter an, welches die Person zum Zeitpunkt der Erfassung hatte. Das Alter wird als Integer angegeben und ist dahingehend diskret z.B. 43 Jahre. Die Verteilung des Datenbereiches, werden im folgendem in diesem Format angegeben. (Quantile [Min, 25, 50, 75, Max]), Quantile [9, 29, 40, 52, 69]

Gender: Das Geschlecht wird als String abgespeichert nimmt aber nur zwei Werte an: "Male" oder "Female". Dabei gibt es einen Anteil von 50 Prozent Männern und 50 Prozent Frauen.

Bedtime: Gibt die Uhrzeit an zu der die Person ins Bett gegangen ist hierbei ist nicht klar ob damit der Zeitpunkt gemeint ist, zu dem die Person eingeschlafen ist oder zu dem die Person

sich ins Bett gelegt hat. Die Information wird als DateTime angegeben. Die Daten steigen in 30 Minuten schritten und ist trotz DateTime somit Diskret.

WakeUp Time = Gibt das Datum und die Uhrzeit an zu dem die Person erwacht steigt analog zu der Bedtime in halben Stunden Schritten an. Das Datum ist bei beiden DateTime formaten nicht von weiterem interesse, da es keine zeitliche Entwicklung der erfassten Personen gibt.

Sleep Duration: Die Schlafdauer ist wie die Bed Time unklar in Ihrer interpretation, da sich der Wert immer aus der Differenz zwischen bedtime und WakeupTime berechnet. Daher ist unklar ob es sich um die geschlafene oder um die im Bett verbrachte Zeit handelt. Die Spalte wird als Float angegeben und steigt aufgrund der halbstündlichen Sprünge der BedTime und Wakeup-Time auch in 0.5 Schritten. Die Daten haben Quantile von [5.0, 7.0, 7.5, 8.0, 10.0] Stunden. Die Interpretation der Sleep Duration wird weiter dadurch erschwert, dass im weiteren Datensatz die Anzahl angegeben wird wie oft eine Person in der Nacht wach wird. Aber ohne Angabe wie lange diese Schlafpausen spezifisch sind.

Schlaf Effizienz = Gibt den prozentualen Anteil an, die eine Person Schlafend im Bett verbracht hat. Die Daten werden als Float mit zwei Nachkommastellen angegeben. Die Daten haben die Quantile: [0.5, 0.7, 0.82, 0.9, 0.99]. Eine Person die 5 Stunden im Bett verbracht hat und davon eine Stunde wach war. Hat also eine Schlaffeffizienz von 80 Prozent. Da hier wieder die Interpretationsproblematik besteht. Wird im weiteren davon ausgegangen, dass die Schlaffeffizienz angibt welchen Anteil die Person nach dem Einschlafen schlafend, also in einem der drei Schlafzyklen verbracht hat.

REM Sleep percentage = Die REM steht für Rapid Eye Movement Schlaf, dies ist einer der drei Schlafzyklen die ein Mensch im Schlaf durchführt. Die CDC empfiehlt einen Anteil von 25 Prozent Healthline sollte. Der REM Percentage gibt den prozentualen Anteil an den die schlafende Person im REM Verbracht hat. Also Anteil REM an Schlaff Effizienz. Die Daten werden als Integer abgespeichert und haben Quantile von [15, 20, 22, 25 30]

Deep sleep percentage? = Der Tiefschlaf Prozentsatz gibt den Anteil am Schlaf an, der im Tiefschlaf verbracht wurde. Die Daten werden als Integer angegeben und haben Quantile: [18, 51, 58, 63, 75]

Light sleep percentage = Gibt den prozentualen Anteil am Schlaf an, der im Leichtschlaf verbracht wurde. Die Daten werden als Integer angegeben und haben Quantile von [7, 15, 18, 40, 63]

Awakenings = Gibt die absolute Anzahl an, wie oft eine Person aufgewacht ist. Die Daten werden im Datensatz als Float abgespeichert. 0.0 bedeutet eine Person hat durchgeschlafen und ist nur einmal Final am morgen aufgewacht. Die Daten reichen von [0.0, 1.0, 1.0, 3.0, 4.0]

Caffeine Intake = Gibt an wie viel Koffein die Person in den letzten 24 Stunden zu sich genommen hat. Die Maßeinheit hierbei beträgt mg. Die Daten werden als Float abgespeichert und haben Quantile von [0.0, 0.0, 25.0, 50.0, 200.0].

Alcohol Intake = Gibt an wie viel Alkohol die Personen in den letzten 24 Stunden zu sich genommen hat in Oz. Die Daten haben Quantile von [0.0, 0.0, 0.0, 2.0, 5.0]

Tobacco Intake = Gibt an ob die Person Raucht. Die Daten sind als String abgespeichert: Yes für Raucher und No für Nichtgeraucht. 154 Personen geben an zu Rauchen und 298 geben an NichtRaucher zu sein.

Exercise Intake = Gibt an wie viele Einheiten Sport die Person in der Woche macht. Dabei ist nicht angegeben welche Maßeinheit diese Einheiten Sport haben. Die Daten haben Quantile von [0, 0, 2, 3, 5], es gibt 6 fehlende Werte.

**Bei Annahme, dass die Daten legitim sind und die angegebene Interpretation der Attribute korrekt ist, ermöglicht dieser Datensatz neue Einblicke in die Schlafqualität vor allem die Schlafphasen und Dauer dieser. Zusätzlich werden relevante Verhaltensweisen und Einflüsse erfasst. Die Erfassung der Einnahme von Kaffee und Alkohol ist, suboptimal, da die beiden Substanzen innerhalb des im menschlichen Körpers eine geringe Halbwertszeit aufweisen. Dahingehend wäre der Zeitpunkt der Einnahme relevant. Weitere Faktoren die die Schlafqualität beeinflussen werden nicht erfasst weiter werden die Länge der individuellen Schlafunterbrechungen nicht differenziert und den drei Schlafphasen nicht zugeordnet. So könnte es für den Anwender von Interesse zu sein, welche der Schlafphasen durch welche Verhaltensweisen gestört werden. Weiter lässt sich argumentieren, dass die Verteilung der Verhaltensweisen sehr linksseitig ist. Dies mag auf kulturelle Unterschiede zurückzuführen sein. Daher muss die Aussagekräftigkeit des Datensatzes auf den Marrokanischen Lebensstil eingeschränkt werden.**

**Die Daten** Der Datensatz wurde auf Kaggle veröffentlicht und unterliegt keinem Copyright Schutz. Die Daten sind als CSV mit einer Größe von 9kB zugänglich. Die Daten sind in dem Github Repository abgespeichert, dies ermöglicht eine zentrale Aktualisierung der Daten, falls dies nötig sein sollte. Das Programm ruft die CSV Datei mittels eines Http.get request auf und speichert diese als einen String ab. Falls der request klappt wird die Message 'Got Text Ok fulltext' an die Funktion 'update' gesendet. Der String ist hierbei repräsentiert durch 'fullText'. Daraufhin wird im 'Model' das 'datenladen' auf Success gesetzt. Auf den String wird die Funktion 'stringtoUnverarbeitete', deren Ziel es ist den String in eine List(UnverarbeiteteDaten) zu transformieren. Die Funktion benutzt das Paket BrianHicks/elm-csv (Im Code als Decode) Diese zieht die Namen der Felder aus der ersten Reihe in dem String. Die Funktion decode decodiert die Inputdaten, relevant hierbei ist das, auch leere Felder in dem String vorkommen dürfen. Die Funktion Decode.blank gibt ein 'Nothing' zurück, falls das Feld leer sein sollte. Wenn es doch zu einem Fehler kommen sollte gibt die Funktion stringtoUnverarbeitete eine leere Liste zurück an das Modell. Wenn die Daten nun in der Form Unverarbeitete Daten sind wird die Funktion 'sleep2Point' angewandt. Diese entfernt einen Tupel, wenn eines seiner Felder ein 'Nothing' beinhaltet. Weiter werden die Werte für REM, Tiefschlaf und Leitschlaf in wirklich in diesen verbrachten Stunden transformiert indem diese, mit 0.1 in Prozente übertragen wurden und dann mit der Schlaffeffizienz und Schlafdauer multipliziert. Dies ist besser, da somit die Interpretation erleichtert wird, und die Vergleichbarkeit wird hergestellt. Vor der Transformation waren

die Schlafphasen einer Person die 5 Stunden schläft und einer Person die 10 Stunden schläft nicht unterscheidbar. Jetzt lässt sich klar anzeigen wie viel die Personen in den Schlafphasen verbracht haben. Gender und Raucher werden von einem String zu einem Float mittels case handling konvertiert (genderToFloat und raucherToFloat). Dies ermöglicht es mit den Daten zu rechnen und erleichtert das weiter visualisieren. Die Funktion sleep2Point hat einen Output von Typ Alias 'Aussortierte Daten'. Weiter werden die Attribute Bedtime und Wakeuptime in den aktuellen Visualisierungen nicht benutzt, könnte es bei der Weiterentwicklung interessant werden, daher wurden Sie im Datenkonstrukt belassen aber nicht weiter behandelt. Man könnte von einer sanften Projektion sprechen. Innerhalb der Einstellungen kann der Anwender eine Selektion durchführen und Einschränkungen auf das zu untersuchende Attribut anwenden. So werden nur Datentupel an die Visualisierungen übergeben die dem Kriterium entsprechen.

## 3 Visualisierungen

### 3.1 Analyse der Anwendungsaufgaben

Die Aufgaben die durch die Visualisierung gelöst werden sollen:

Die Hauptaufgabe der Visualisierungen ist es dem Nutzer dabei behilflich zu sein sich einen Überblick über die gegebenen Daten zu machen, und die Attribute auf ihren Zusammenhang zu anderen zu analysieren und dann die Größe der Einflüsse der Verhaltensweisen auf das untersuchte Attribut zu differenzieren. Der Anwender muss zunächst die Datenlage kritisch analysieren und die Güte der einzelnen Attribute auf ihre Aussagekräftigkeit zu bewerten. Nach Bewertung der Attribute vergleicht er diese mit anderen Attributen und kann so komplexe Zusammenhänge und Beziehungen erkennen. Daraufhin müssen die Einflüsse des Verhaltens auf die Attribute analysiert werden. Dabei muss der Anwender die Grundlagen der statistischen Interpretation und Relevanz von Normalverteilungen für die Interpretation von Datenlagen kennen. Er sollte also wissen, dass die Datenlage normalverteilt sein sollte, um statistische Aussagen treffen zu können. Weiter ist es von Nutzen die Korrelation interpretieren zu können. Die Interpretation von Rot als negative und Grün als positive ist von Nutzen um die Ergebnisse schneller zu erfassen, ist allerdings nicht notwendig.

Das Problem ist folgendes: Daher ist es wichtig, dass wir die Daten visualisieren um die Zusammenhänge zu erkennen. Die Visualisierung soll uns also erlauben den Einfluss von multiplen Verhaltensindikatoren auf die Ausprägung eines Merkmales zu schätzen.

Analysieren sie die konkreten Anwendungsaufgaben, die die Lösung des Zielproblems durch die Anwender:innen bearbeitet werden müssen.

Welche sinnvollen mentale Modelle helfen den Personen bei der Bearbeitung. Aufgabenstellung: Analyse der Variablen und den Einfluss der Verhaltensindikatoren. Handelt sich um explorative Visualisierung. Mentale Modelle: Was ist eine Normalverteilung? Abweichungen von der Linie: Ein Verständnis dafür, wie Abweichungen von der diagonalen Linie im QQ-Plot auf nicht-normale Verteilungen oder systematische Abweichungen hinweisen können. Schwänze und Spitzen: Das mentale Modell von Schwänzen und Spitzen in einer Verteilung hilft dabei zu verstehen, wie sich Ausreißer oder starke Konzentrationen von Werten auf den QQ-Plot auswirken können.

Sind diese mentalen Modelle für sie notwendig, um die Aufgaben lösen zu können? Gehen sie bei ihrer Argumentation von den Anwendungsaufgaben aus und kommen sie dann zu den mentalen Modellen, deren Aufbau durch Visualisierungen unterstützt wird.

## 3.2 Anforderungen an die Visualisierungen

Leiten sie Anforderungen an das Design der Visualisierungen ab, die sich durch ihre Analyse des Zielpblems ergeben.

**Anforderungen an das Design.**

## 3.3 Präsentation der Visualisierungen

### 3.3.1 Visualisierung Eins: QQ-Plot Datenlage Normalverteilt?

Die erste Visualisierung ist die Darstellung der Daten in einem Scatterplott. Spezifisch der eines Normalverteilung QQ Plots. Ein QQ-Plot (Quantile-Quantile-Plot) ist ein nützliches grafisches Werkzeug in der statistischen Datenanalyse, insbesondere wenn es darum geht, die Verteilung von Daten zu überprüfen und mit einer theoretischen Verteilung zu vergleichen, in der Anwendung wird, Aufgrund des Implementationaufwandes, nur der Vergleich mit der Normalverteilung vorgenommen. Die Visualisierung zeigt den Wert auf der Y Achse an, den ein Datenpunkt hätte, falls der Datensatz einer Normalverteilung folgen würde. Diese werden dann auf einem zwei dimensionalem Diagramm dargestellt. Die X Achse zeigt die Quantile der Daten an, die Y Achse zeigt die Quantile der Normalverteilung an.

Die Typischen Aufgabenfelder eines QQ Plots sind die Normalitätsüberprüfung, die Identifikation von Abweichungen, die Bestätigung der Normalverteilungshypothese und Identifikation von Ausreißern. Bei der Normalitätsüberprüfung, der primäre Zweck eines QQ-Plots, geht es darum zu überprüfen ob ein Attribut einer Normalverteilung folgt. Ein perfekter QQ-Plot würde alle Punkte auf der Linie der Gleichverteilung platzieren. Abweichungen von dieser Linie deuten auf Abweichungen von der Normalverteilung hin. Der Anwender kann in der Visualisierung sehen ob die Attribute einer Normalverteilung folgen. Falls diese Perfekt Normalverteilt sein sollten, liegen die Datenpunkte direkt auf der schwarzen Linie. Die in der Grafik eingezeichnete rote Line ist vom 25 Quantil bis zum 75 Quantil gezogen. Dies lässt den Anwender überprüfen ob es eine additive Verschiebung gibt. Die Identifikation von Ausreißern, können im Normal QQ-Plot, an den visuellen Enden identifiziert werden. Starke Abweichungen sind durch eine große Distanz zur schwarzen Linie identifizierbar. Diese können durch das Filtern des Datenbereiches entfernt werden. weiter ist es möglich einen Vergleich verschiedener Verteilungen anzustellen. Dies ist aufgrund des Umfang des Projektes nicht in der Anwendung implementiert Dies ermöglicht es dem Anwender die Daten und Interpretierbarkeit einzuschätzen und diese in geringem Maße mit den Filtern zu Selektieren. Dies erlaubt es dem Anwender die aussagekräftigen Attribute von den aussageschwachen zu unterscheiden.

Das oben Auswählen und das Filtern geschieht in der Seitenleiste, hier kann der Benutzer das Attribut auswählen, dass er untersuchen möchte und direkt darunter dieses in seinem Wertebereich filtern. Der eingeschränkte Datenbereich wird dann in den anderen Visualisierungen verwendet. Der Benutzer kann in der zweiten Visualisierung zwei weitere Attribute zum untersuchen auswählen. Diese werden daraufhin unter dem Hauptscatterplott als kleinere QQ-Plott



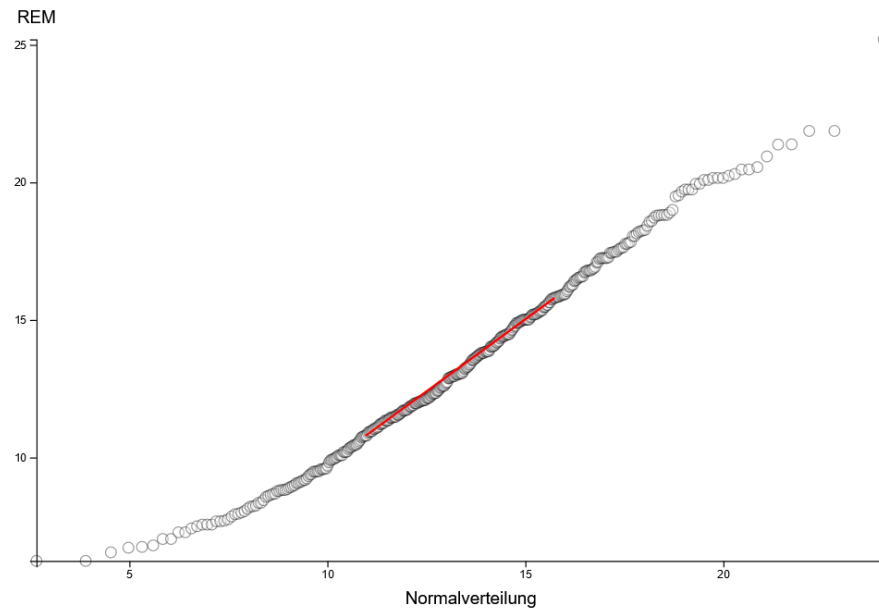


Abbildung 1: Visualisierung Eins Norm QQ-Plot REM

dargestellt und ermöglichen eine schnelle Einschätzung der Güte. Hier mit wird visualisiert, dass die Attribute mit denen verglichen wird den gleichen Ansprüchen gegenüberstehen muss wie das Hauptattribut.

Alternativ hätte auch ein Histogramm verwendet werden können, bei diesem hätte ein Distributionfitting leichter stattfinden können.

Expressivität und Effektivität der Visualisierung diskutieren.

Die Zielgruppe der Visualisierungen sind Nutzer mit einem Interesse an den zu grundlegenden Daten.

### 3.3.2 Visualisierung Zwei: Parallele Koordinaten

Präsentieren der Visualisierung und Interaktionsmöglichkeiten. 3 Dimensionen werden abgebildet. Die in QQ-Plott ausgewählte sowie zwei weitere Attribute. Die Achsen sind parallel zueinander angelegt und gleich groß. Die Ausprägungen eines Datupels werden dargestellt, indem man die Ausprägung des der Tupel am Attribut hat auf die Y-Achse skaliert. Dann wird eine gerade Linie zwischen den Positionen auf den anliegenden Achsen gezogen. Die Endpunkte der Liniensegmente korrespondieren also mit den Attributwerten. Das Hauptattribut, welches in Visualisierung 1. ausgewählt wurde ist in der Mitte angelegt. Das Y-Attribut zu der linken und das Z-Attribut zu der rechten Seite. Es wurde bewusst darauf verzichtet weitere Attributdimensionen hinzuzufügen, da es einerseits der Anwender einen Fokus auf die Analyse des

Hauptattributen hat und andererseits, bei mehr als 3 Attributen zusammenhänge suggeriert werden könnten, die nicht unbedingt vorhanden sein müssen. Nur weil eine Clusterung zwischen Dimension 1 und 2 und zwischen 2 und 3 besetzt. Beduetet das nicht immer, dass es eine Clusterung zwischen 1 und 3 gibt. Dies ist der Effektivität abträglich, da so zusammenhänge suggeriert werden könnten die nicht intendiert sind.

Die Parallelen Koordinaten eignen sich gut für mehrdimensionale, quantitative und endliche Daten. (Dimension < 20)

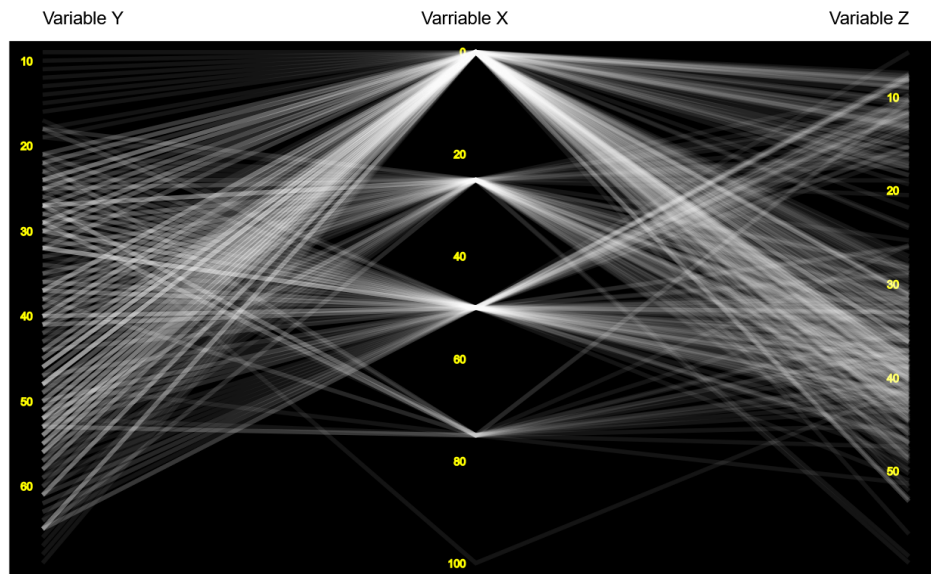


Abbildung 2: Visualisierung Eins Norm QQ-Plot REM

Wieso sind diese gut und erfüllen die Anforderungen an das Design?

Wieso erfüllt die Visualisierung den Anforderungen der Anwender?

Der Anwender hat ein Interesse die Zusammenhänge zwischen dem Hauptattribut und den anderen Attributen zu untersuchen, die nicht leicht mit statistischen Test erkannt werden können. Durch die Betrachtung der Linien können Muster, Trends oder Gruppierungen in den Daten identifiziert werden, die auf Beziehungen zwischen den attributen Hinweisen. Zum Beispiel können parallele Linien auf eine starke Korrelation zwischen den entsprechenden Variablen hinweisen. Der Anwender kann somit Rückschlüsse auf den zusammenhang von Attributen ausbilden und diese mithilfe von Visualisierung NR. 3 Rückschlüsse auf die, hinterden Daten liegenden Mechanismen schließen und somit sein Verständnis von Schlafqualität zu verbessern. Die Untersuchung der Attribute und das Visualisieren kann mittels eines interaktiven anpassens der Pixelgröße und RGBA Werte erfolgen, dies ist vorallem nützlich, da viele der Inputdaten wenige unterschiedliche Wertausprägungen haben, was die Clusterung der Inputdaten erschwert. Um dennoch interessante Muster erkennen kann. So kann der Benutzer basierend auf seinen eigenen visuellen Bedürfnissen anpassungen vornehmen. In Bild 3 wird dies vorgestellt.

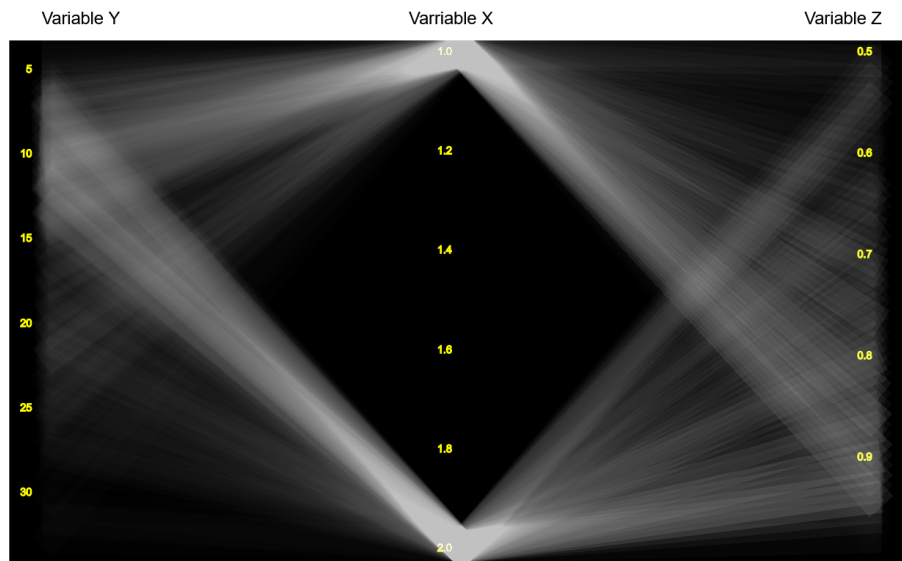


Abbildung 3: Visualisierung Zwei: Parallele Koordinaten:Pixel 30, RGBA (255, 255, 255, 0.006)

Wahrnehmungsprinzipien und Theorie über Informationsvisualisierung verweisen: Clusterin, Farbwahrnehmung und Abstrahierung

Wieso diese Visualisierung und nicht andere? Wir wollen echte mögliche Alternativen die zu einem ähnlichem mentalem Modell führen:

Streuudiagramm Matrix,

Expressivität und Effektivität der Visualisierung diskutieren.

### 3.3.3 Visualisierung Drei: Force Graph

Test

Präsentieren der Visualisierung und Interaktionsmöglichkeiten. Wieso sind diese gut und erfüllen die Anforderungen an das Design?

Wieso erfüllt die eVisualisierung den Anforderungen der Anwender?

Wahrnehmungsprinzipien und Theorie über Informationsvisualisierung verweisen: Wir bauen ein Kausales Mentales Modell auf, Diese verhaltensweisen beeinflussen dieses Attribut in dieser Art und Weise.

Wieso diese Visualisierung und nicht andere? Wir wollen echte mögliche Alternativen die zu einem ähnlichem mentalem Modell führen.

Expressivität und Effektivität der Visualisierung diskutieren.

### Kausales Mentales Modell

### 3.4 Interaktion

Die präsentierten Visualisierungstechniken müssen interaktiv zu einer Anwendung verknüpft werden.

1 erste Visualisierung: Auswahl des Attributs -> Hauptattribut der beiden anderen Visualisierungen Filtern des Datenbereiches. -> Schränkt die Datenmenge ein auf die sich die anderen beziehen.

2. zweite Visualisierung: Auswahl der Attribute -> In 1 werden Junior QQ plots aufgerufen. -> Und an die SSchneeflocke übergeben.

Die Interaktion mit einer Visualisierung soll in den anderen Visualisierungen zu einer Änderung führen. Erklären Sie die möglichen Interaktionen mit den einzelnen Visualisierungen und die möglichen Verknüpfungen zwischen ihnen. Begründen Sie warum die konkreten Interaktionen umgesetzt wurden und welche Zwecke für die Anwenderinnen mit ihnen unterstützt werden. Begründen Sie ebenfalls warum Sie andere Interaktionsmöglichkeiten nicht umgesetzt haben. Wenn Sie keine der geforderten Interaktionen umsetzen, erhalten Sie im gesamten Projekt deutlichen Punktabzug.

## 4 Implementierung

Beschreiben Sie die Implementierung ihrer Visualisierungsanwendung in Elm. Onepage application, da die Interaktion mit den Filtern und Darstellungen einfach sein soll. Links Einstellungen und Filter, daneben die Visualisierungen mit Erklärungen. Die Visualisierungen werden im 'view' als Funktionen aufgerufen, die ein Svg msg produzieren, welches abgebildet wird. Gespeichert und dann in der View Funktion aufgerufen.

Stellen Sie die Gliederung Ihres Quellcodes vor.:

Der Code besteht aus einem Main Modul, welches die anderen Module importiert, die Daten einlädt und die Interaktionen zwischenspeichert. Die drei Visualisierungen sind in jeweils einem anderem Modul implementiert, dabei werden die benötigten Informationen übergeben.

Das Modell speichert die Ausprägungen ab, den Zustand des Einladens, die Wahl der Attribute in den Dropdowns, die Daten werden als Liste aussortierter Daten abgespeichert.

Die Filter und die Pixelgröße und die RGBA Werte werden als einfache Float abgespeichert.

Da Elm keine Rückwärtskompatibilität erzwingt, ist es manchmal nicht möglich die neueste Version eines Paketes zu verwenden. Dies ist allerdings etwas frustrierend, wenn man ein Nützliches Paket gefunden hat. Es ist allerdings möglich den Quellcode auf Github, lokal abzuspeichern und als anderes Paket zu importieren, solange die Abhängigkeiten erfüllt sind. Das Modul JXStat ist das Paket Stat aus dem Elm-Stat 6.0.2.

Haben Sie verschiedene Elm-Module erstellt. Was war aufwändig umzusetzen, was ließ sich mit dem vorhandenen Code aus den Übungen relativ einfach umsetzen?

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

Die Interaktionen werden im Modell als einfache String oder Floats abgespeichert. Der Datensatz wird also nach seiner Transformierung in der Init nicht weiter verarbeitet.

Da die Interaktion ist vor allem durch Projektion und Selektion umgesetzt, dies ist in der Anwendung durch die Filter und die Auswahl der Attribute umgesetzt. Die Anpassung der Daten an die Anforderungen der Visualisierungen erfolgt im Main vor der Übergabe in die Grafiken.

## 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 den Aufwand und die Schwierigkeiten Ihres Ansatzes und der Alternativen.

### 5.1 Anwendung Visualisierung Eins

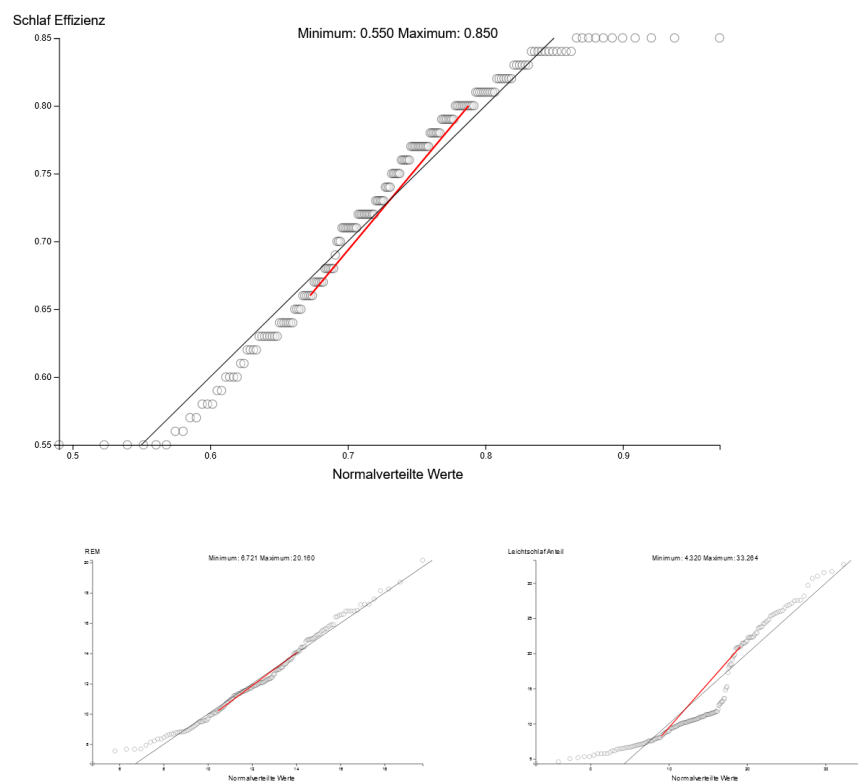


Abbildung 4: Visualisierung Eins: Anwendungsbeispiel

## 5.2 Anwendung Visualisierung Zwei

Im zweiten Schritt der Visualisierung wurden die Attribute REM und Leichtschlaf Anteil untersucht. Man kann gut erkennen, dass es einen Zusammenhang zwischen niedrigem Leichtschlaf und hoher Schlafeffizienz gibt. Bei hohem Leichtschlaf Anteil geht ein leichtes Cluster zu einer reduzierten Schlafeffizienz. Der Zusammenhang zwischen REM und Schlafeffizienz besteht aus parallel absinkenden Strichen, dies deutet darauf hin, dass mit Anstieg des REM Schlafanteils die Schlafeffizienz abnimmt.

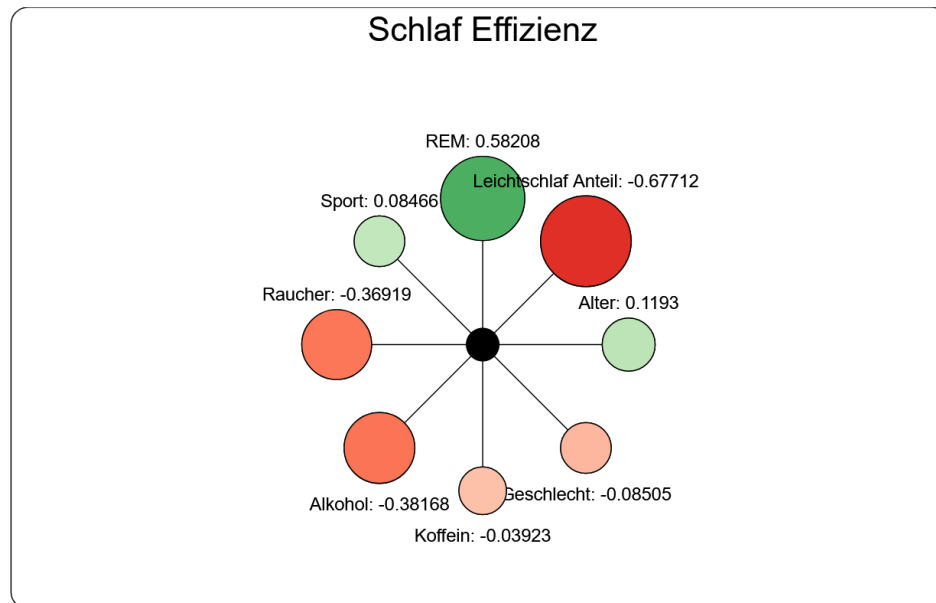


Abbildung 5: Visualisierung Eins: Anwendungsbeispiel

So lässt sich nur aus der Visualisierung ohne Vorwissen die Hypothese entwickeln, dass mit geringerer Schlafeffizienz die REM Schlafanteile sinken und der Anteil des leichten Schlafes erhöht wird.

## 5.3 Anwendung Visualisierung Drei

Die durch Visualisierung 2 entwickelte Hypothese, kann durch Visualisierung 3 unterstützt werden. Die Repräsentation der Korrelation durch die Kugeln in Größe und Farbe zeigen unterstützen, dass die Korrelation von REM auf die Schlafeffizienz positiv ist und die Korrelation des Leichtschlafanteils negativ. Weiter kann man erkennen, dass die Standardmäßig angezeigten Verhaltensweisen Attribute mit der Schlafeffizienz korrelieren.

## **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**

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**