

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

1.2 Zielgruppen

[2]

Die Schlafqualität spielt eine entscheidende Rolle für die Lebensqualität, deswegen richtet sich die Anwendung an eine Zielgruppe, die ein Interesse daran hat, die Zusammenhänge und Einflüsse auf die Schlafqualität zu verstehen. Diese Zielgruppe könnte aus Forschern und Privatpersonen bestehen, die daran interessiert sind, wie verschiedene Lebensstile die Schlafqualität beeinflussen können und die Schlafzyklen sich zueinander verhalten. Ein tiefgehendes Wissen über die Schlafphasen ist nicht erforderlich, aber ein generelles Interesse an der Analyse des Zusammenhangs zwischen Verhalten und Schlafqualität charakterisiert den Benutzer. Die Benutzer sollten sich darüber im Klaren sein, dass die Visualisierungen Hinweise liefern können, jedoch nicht notwendigerweise kausale Zusammenhänge zeigen.

Die Visualisierungen, die in diesem Projekt präsentiert werden, sollen dabei helfen, Einflüsse und Zusammenhänge bei der Schlafqualität zu identifizieren. Dabei werden verschiedene Attribute wie Koffein-, Alkohol- und Tabakkonsum, Sport, sowie Alter und Geschlecht der Personen untersucht. Die Zielgruppe sollte ein Grundverständnis für die Bedeutung dieser Attribute in Bezug auf die Schlafqualität haben.

Die Benutzer sollten sich bewusst sein, dass der Datensatz, der in dieser Anwendung verwendet wird, nur 452 Personen umfasst und durch die Datenverarbeitung weiter reduziert wird. Daher ist es wichtig, die statistische Validität der Ergebnisse zu berücksichtigen, insbesondere bei der Anwendung von Filtern. Statistische Aussagekraft ist bei $n < 30$ nicht mehr gegeben, und daher müssen die Benutzer dies bei der Benutzung berücksichtigen.

1.3 Überblick und Beiträge

Im diesem Abschnitt wird ein Überblick auf die Daten und verwendeten Visualisierungstechniken gegeben.

2 Daten

Der verwendete Datensatz bietet eine umfangreiche Sammlung von 15 Attributen, die Informationen über die Schlafqualität und Lebensgewohnheiten von 452 Personen liefern. Auf der Kaggle Seite wird nach Angaben des Authors behauptet, dass der Datensatz im Kontext einer Studie von der ENSIAS, Marroco gesammelt wurde. Innerhalb einer Literatur Recherche konnten weder auf der Webseite der ENSIAS noch in weitergehender Literaturrecherche eine Veröffentlichung zu diesem Datensatz identifizieren. Daher sollten die Daten und daraus entwickelten Ergebnisse, nicht unreflektiert übernommen werden.

Der Datensatz hat drei Nominale (Id, Gender und Raucher) und zwölf Quantitative (Age, Bedtime, Wakeup Time, Sleep Duration, Sleep Efficiency, REM Sleep percentage, Deep Sleep Percentage, Light Sleep Percentage, Awakenings, Caffeine Intake, Alkohol Intake, Tobako Intake, Exercise frequency) Aufgrund der unklaren Herrkunft des Datensatzes muss werden die Attribute im folgenden näher beschrieben.

- ID: Die ID identifiziert eine Person eindeutig. Da keine ID mehrfach aufgeführt wurde, ist anzunehmen, dass jede Person nur einmal an der Studie teilgenommen hat. Die ID wird als Integer bereitgestellt, und der Datenbereich reicht von 1 bis 452.
- Age: Gibt das Alter an, das die Person zum Zeitpunkt der Erfassung hatte. Das Alter wird als Integer angegeben und ist daher diskret, zum Beispiel 43 Jahre. Die Verteilung des Datenbereichs wird im Folgenden in diesem Format angegeben: (Quantile [Min, 25, 50, 75, Max]), Quantile [9, 29, 40, 52, 69].
- Gender: Das Geschlecht wird als String abgespeichert und nimmt 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 werden als DateTime angegeben. Die Daten steigen in 30 Minuten schritten und ist trotz DateTime Formati damit 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 Bedtime 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 WakeupTime 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, ohne Angabe, wie lange diese SSchlafpausen spezifisch sind.

- Sleep efficiency: Gibt den prozentualen Anteil an, den 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 Schlafeffizienz von 80 Prozent. Da hier wieder eine Interpretationsproblematik besteht, wird im weiteren davon ausgegangen, dass die Schlafeffizienz angibt, welchen Anteil die Person nach erstmaligem Einschlafen, schlafend verbringt. Weiter wird diese Kennzahl in der Datenverarbeitung auf einen absoluten Stundenwert hochgerechnet.
- 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 . Der REM Percentage gibt den prozentualen Anteil an, den die schlafende Person im REM verbracht hat, also den Anteil REM an Schlaf Effizienz. Die Daten werden als Integer abgespeichert und haben Quantile von [15, 20, 22, 25, 30]. Weiter wird diese Kennzahl in der Datenverarbeitung auf einen absoluten Stundenwert hochgerechnet.
- 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 von [18, 51, 58, 63, 75]. Weiter wird diese Kennzahl in der Datenverarbeitung auf einen absoluten Stundenwert hochgerechnet.
- 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]. Weiter wird diese Kennzahl in der Datenverarbeitung auf einen absoluten Stundenwert hochgerechnet.
- Awakenings: Gibt die absolute Anzahl an, wie oft eine Person aufgewacht ist. Die Daten werden im Datensatz als Float abgespeichert. 0.0 bedeutet, dass eine Person durchgeschlafen und nur einmal final am Morgen aufgewacht ist. 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 haben 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 Nichtraucher. 154 Personen geben an zu rauchen und 298 geben an Nichtraucher zu sein.
- Exercise frequency: 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.

Unter der Annahme, dass die Daten valide sind und die angegebene Interpretation der Attribute korrekt ist, gewährt dieser Datensatz detaillierte Einblicke in die Schlafqualität, insbesondere hinsichtlich der Schlafphasen und ihrer Dauer. Allerdings weist die Erfassung der Kaffee- und Alkoholeinnahme Schwächen auf, da diese Substanzen im menschlichen Körper eine geringe Halbwertszeit haben, und somit der zeitliche Kontext der Einnahme relevant wäre. Weitere Einflussfaktoren auf die Schlafqualität werden nicht erfasst, ebenso wenig wird die Länge individueller Schlafunterbrechungen differenziert oder den drei Schlafphasen zugeordnet. Daher könnte es für den Anwender von Interesse sein zu erfahren, welche der Schlafphasen durch bestimmte Verhaltensweisen gestört werden. Zudem lässt sich argumentieren, dass die Verteilung der Verhaltensweisen sehr linksseitig ist, dies ist möglicherweise auf kulturelle Unterschiede zurückzuführen. Daher sollte die Aussagekraft des Datensatzes auf den marokkanischen Lebensstil eingeschränkt werden. **Die Daten** Der Datensatz wurde auf Kaggle veröffentlicht und ist die dritte Version. Die Daten sind als CSV mit einer Größe von 9 kB zugänglich. Sie sind im GitHub-Repository abgelegt, was eine zentrale Aktualisierung der Daten ermöglicht, falls dies notwendig sein sollte.

Das Programm ruft die CSV-Datei mittels eines HTTP-GET-Requests auf und speichert sie als einen String ab. Falls der Request erfolgreich ist, wird die Nachricht 'Got Text Ok fulltext' an die Funktion 'update' gesendet. Der String wird dabei durch 'fullText' repräsentiert. Daraufhin wird im 'Model' das 'datenladen' auf 'Success' gesetzt. Auf den String wird die Funktion 'stringtoUnverarbeitete' angewandt, deren Ziel es ist, den String in eine 'Liste (UnverarbeiteteDaten)' zu transformieren. Die Funktion verwendet das Paket 'BrianHicks/elm-csv' (im Code als Decode bezeichnet), um die Namen der Felder aus der ersten Zeile im String zu extrahieren. Die Funktion 'decode' decodiert die Eingabedaten, wobei zu beachten ist, dass auch leere Felder im String vorkommen dürfen. Die Funktion 'Decode.blank' gibt ein 'Nothing' zurück, falls das Feld leer ist. Wenn es dennoch zu einem Fehler kommt, gibt die Funktion 'stringtoUnverarbeitete' eine leere Liste an das Modell zurück. Wenn die Daten nun in Form von 'UnverarbeiteteDaten' vorliegen, wird die Funktion 'sleep2Point' angewandt. Diese entfernt ein Tupel, wenn eines seiner Felder ein 'Nothing' enthält. Weiter werden die Werte für REM, Tiefschlaf und Leichtschlaf in tatsächlich verbrachte Stunden transformiert, indem sie mit 0.1 in Prozent umgerechnet und dann mit der Schlafeffizienz und Schlafdauer multipliziert werden. Dies verbessert die Interpretierbarkeit und

stellt die Vergleichbarkeit her. Gender und Raucher werden von einem String zu einem Float mittels Case-Handling konvertiert (Funktionen ‘genderToFloat’ und ‘raucherToFloat’). Dies ermöglicht das Rechnen mit den Daten und erleichtert die weitere Visualisierung. Die Funktion ‘sleep2Point’ hat einen Output vom Typ Alias ‘AussortierteDaten’. Weiter werden die Attribute Bedtime und Wakeup Time in den aktuellen Visualisierungen nicht verwendet, könnten jedoch 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 konkreten Anwendungsaufgaben, die von den Anwendern bearbeitet werden müssen, zielen darauf ab, die Daten kritisch zu analysieren, Zusammenhänge zwischen verschiedenen Attributen explorativ zu erkennen und den Einfluss der Verhaltensweisen zu erkennen.

1. Datenlage analysieren:

- Die Nutzer müssen zunächst die Datenlage kritisch analysieren und die Güte der einzelnen Attribute bewerten. Hierbei ist ein Verständnis für die statistische Interpretation und die Relevanz von Normalverteilungen zur Aussagefähigkeit der Visualisierungen wichtig.

2. Vergleich von Attributen:

- Nach der Bewertung der Güte der Attribute, soll ein Nutzer den Zusammenhang dieser mit anderen Attributen verstehen. Es soll hierbei auch auf versteckte Zusammenhänge stoßen können und die indirekte Beziehungen erkennen. Die Visualisierungen unterstützen dabei, Muster und Trends in den Daten zu identifizieren.

3. Analyse der Einflüsse des Verhaltens:

- Die Einflüsse des Verhaltens auf die Attribute müssen analysiert werden. Hierbei ist es notwendig, den Einfluss von multiplen Verhaltensindikatoren auf die Ausprägung eines Merkmals zu schätzen.

Bei diesen Aufgabe sind die folgenden mentalen Modelle von Bedeutung:

1. Statistische Interpretation:

- Nutzer sollten Grundlagen der statistischen Interpretation und die Relevanz von Normalverteilungen verstehen. Visualisierungen wie QQ-Plots können dabei helfen, Abweichungen von der Normalverteilung zu identifizieren.

2. Korrelation interpretieren:

- Ein Verständnis für die Interpretation von Korrelationen ist wichtig. Visualisierungen können durch Farbgebung (z. B. Rot für negative und Grün für positive Korrelationen) unterstützen, auch wenn dies nicht zwingend notwendig ist.

3. Parallele Koordinaten Interpretieren:

- Benutzer sollten verstehen, dass jede Achse ein Attribut repräsentiert, und Linien zwischen den Achsen die Beziehung zwischen Werten in verschiedenen Dimensionen anzeigen. Die Interpretation von Parallelenkoordinaten erfordert die Fähigkeit, Muster in den Linienverläufen zu erkennen und so Zusammenhänge zwischen den Attributen zu verstehen.

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

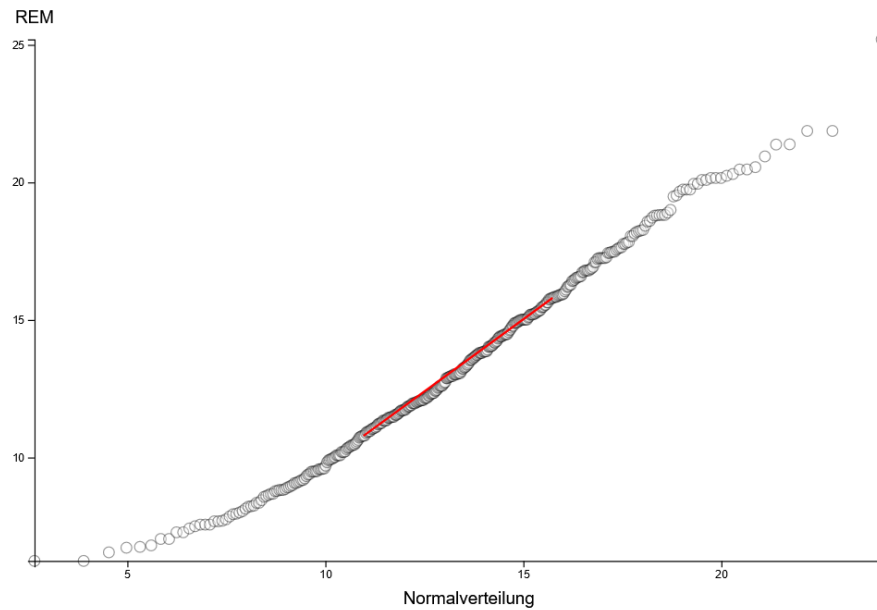


Abbildung 1: Visualisierung Eins Norm QQ-Plot REM

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)

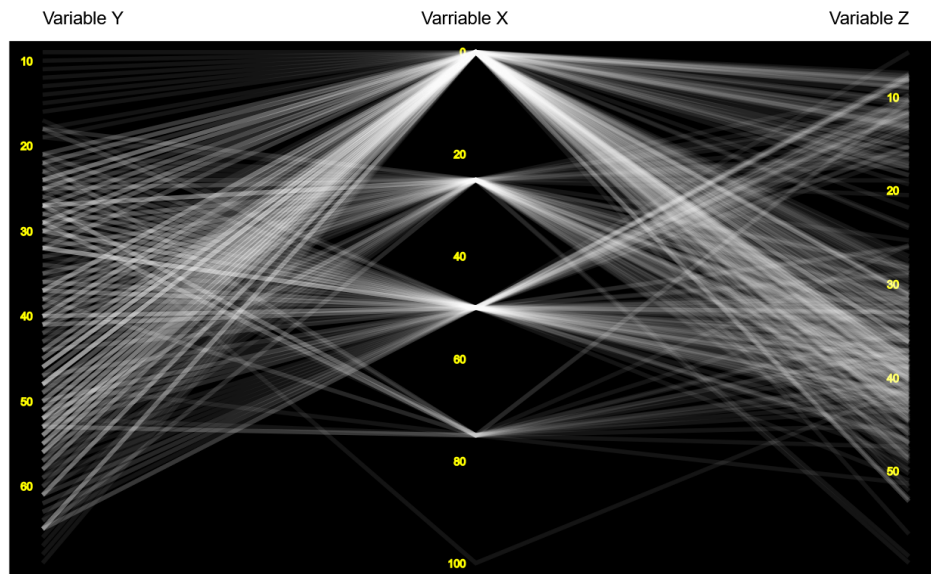


Abbildung 2: Visualisierung Eins Norm QQ-Plot REM

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.

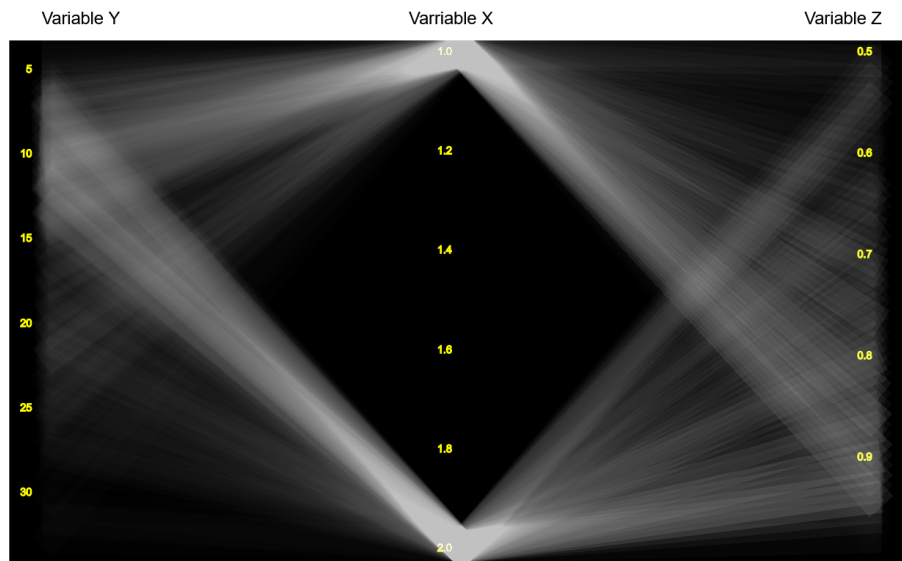


Abbildung 3: Visualisierung Zwei: Parallele Koordinaten: Pixel 30, RGBA (255, 255, 255, 0.006)

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 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 Model besteht aus speichert die Ausprägungen ab, der Zustand des Einladens, die Wahl der Attribute in den Dropdown, die Daten werden als List Aussortierte Daten abgespeichert.

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

Da Elm keine Rückwärts Kompatibilität 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 Model 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

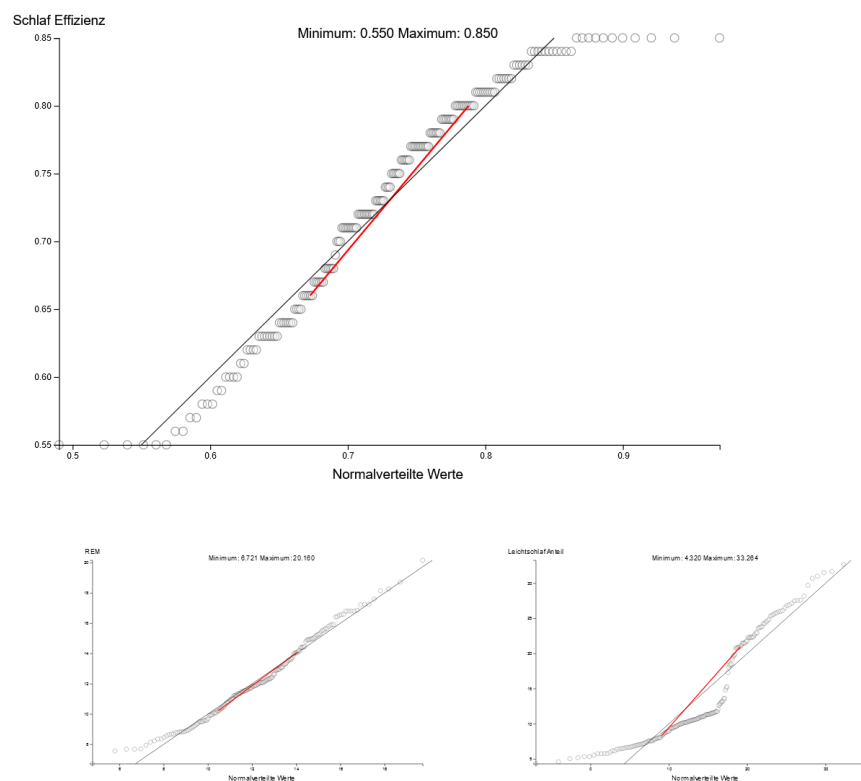


Abbildung 4: Visualisierung Eins: Anwendungsbeispiel

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 Anteil 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.

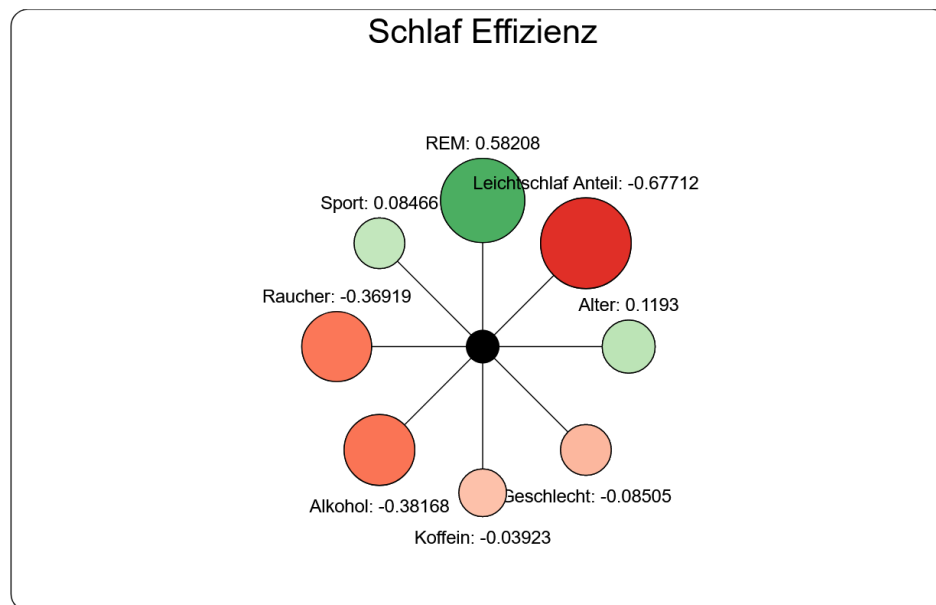


Abbildung 5: Visualisierung Eins: Anwendungsbeispiel

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 stützen, 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