

# Projektbericht zum Modul Information Retrieval und Visualisierung Sommersemester 2023

## Visualisierungen zur Analyse von Einflussfaktoren auf die Schlaf Effizienz

Mick Stewart Wörner

26. 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 [1]: 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 die wissen, dass die Datenlage kritisch Analysiert auf aussagekräftigkeit analysiert werden muss, da der benutze Datensatz mit  $n < XXXX$  relativ 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

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. Input Daten ist csv Datei mit teilweise leeren Datenpunkten. Umgang mit den leeren Feldern-> Mit List.map (Bei einem Nothing, wird der Tupel komplett gelöscht. ), da wir somit nur mit einem sauberem Datensatz arbeiten. Datensatz hat nur Fehlende werte bei gewisse Datenpunkten. (Koffein, Alkohol, Tabak, Sport).

ID : Identifikator der eine Reihe eindeutig Identifiziert. -> Unique identifier keine Person wird zweimal erfasst. -> keine Zeitentwicklung. -> 452 Personen erfasst. Type : Integer.

Age: Gibt das Alter der Person zum zeitpunkt der erfassung hatte.-> Type : Inte-

ger Datenreichweite 9-69 Jahre. Diskreter Wert.

Gender: Gibt das Geschlecht der Person mittels "Male" oder "Female" wieder. Type : String : Boolean Wert.

Bedtime: Gibt die Uhrzeit an zu der die Person ins Bett gegangen ist. Type : DateTime: Diskreter Wert. Halbe Stunde Schritte

WakeUp Time = Gibt das Datum und die Uhrzeit an an dem die Person erwacht ist. Type : DateTime

Sleep Duration = Gibt die Stunden an, die die Person im Bett schlafend verbracht hat in Stunden. Type: Float

Schlaf Effizienz = Gibt die Zeit an die im Bett schlafend verbracht wurde. Type: Float -> In Prozent -> Interpretation nur in combination mit Anzahl erwacht möglich. -> MICK müssen wir das noch Transmutieren? -> Nacher drüber nachdenken.

REM Sleep percentage = Gibt den Prozentualen Anteil an, der im REM Schlaf verbracht wurde. Type: Integer WICHTIG: In Prozent.

Deep sleep percentage? = Gibt den Prozentualen Anteil an, der im Tiefschlaf verbracht wurde. Type: Integer WICHTIG : In Prozent.

Light sleep percentage = Gibt den Prozentualen Anteil an, der im Leichtschlaf verbracht wurde. Type: Integer WICHTIG: In Prozent.

Awakenings = Absolute Anzahl an , wie oft eine Person aufgewacht ist. Interpretation -> 0.0 bedeutet, ungestörter Schlaf. -> Lässt an der Validität des Datensatzes zweifeln. Da, bei "0.0"mal aufgewacht die Person im Schlaf gestorben ist.

Caffeine Intake = Gibt an wie viel Koffein die Person zu sich genommen hat in den letzten 24 Stunden (In mg) ( Frage für Mick: Halbwertszeit von Koffein? (Adenosin Rezeptoren block.) -> Problematik bei der Interpretation. Hoher Koffein intake in the Morning might not have the same effect as the Koffein intake in the evening, Surge of adenosin before going to bed -> More sleep? Or maybe worse sleep?: Datentyp: VERHALTEN ).

Alcohol Intake = Gibt an wie viel Alkohol die Personen in den letzten 24 Stunden zu sich genommen hat. (In OZ) Maßeinheit. Alkohol-> Nerven gift -> Blockierung der REM Tiefschlafphase?

Tobacco Intake = Gibt an wie viel Tabak die Person zu sich genommen hat. -> Boolean Wert. Raucht die Person ja oder nein. (Datensatz wurde angeblich in einem Paper produziert.)

Exercise Intake = Gibt an wie viel Sport die Person getrieben hat. Sporteinheiten pro Woche. (Gibt nur die Absolute Anzahl an, das Problem liegt darin, dass die Daten nicht nach Minute aufgeteilt sind. )

Gehen sie kritisch darauf ein, in wie weit sich die Daten für die Bearbeitung der Fragestellungen und dem Erreichen von Lösungen für die oben beschriebene Zielgruppen eignen. es ist nicht klar ob die Daten legitim sind. Die Quelle auf Kaggle ist nicht angegeben.-> Eignet

sich besonders gut für unseren Zweck, da wir die Datenverteilung und das Verhalten von den Variablen zueinander untersuchen wollen. Haben sie die Daten sinnvoll mit weiteren Datenquellen ergänzt? Wenn ja, wie? Erklären sie die technische Bereitstellung der Daten.

**Daten werden als amerikanisches CSV auf Kaggle.com bereitgestellt. (MICK: Genauen Typ der CSV herausfinden. (trennung mit , statt dem Europäischem ;))**

**Die Daten** Wie sind die Daten zugänglich? Welche Formate werden genutzt. Gibt es Besonderheiten beim Lesen der Formate? Beschreiben sie die Datenvorverarbeitung. Welche Datenvorverarbeitungsschritte sind notwendig? **Einspielen des CSV als String, -> Werden aus dem Github Repository gezogen.** Beschreiben Sie die einzelnen Schritte und begründen sie sie, z.B. warum werden manche Daten weggelassen, über welche Mengen werden Durchschnitte berechnet, warum sind die so berechneten Werte aussagekräftiger als andere Werte. Wenn möglich sollen sie die Datenvorverarbeitung in Elm programmieren, so dass ihre Anwendung auf eine Änderung der Rohdaten reagieren kan.

Realisiert eine Daten-zu-Daten-Abbildung

**Mögliche Operationen:**

- Vervollständigung, Interpolation
- Projektion (Reduzierung der Variablen)
- Selektion (Anwendung von Filterkriterien, Glättung, Ausreißereliminierung)
- Berechnung impliziter Eigenschaften (z.B. Maximum, Gradient)
- Konvertierung

Die Daten sollten so konvertiert werden, dass Sie den Visualisierungsanwendungen entsprechen. Veränderung der Merkmalsausprägungen Daten alle Strings außer ID in einen Interpretierbaren Wert. Unsere Strings sind Boolean, male oder female und Raucher Yes und No

## 3 Visualisierungen

### 3.1 Analyse der Anwendungsaufgaben

Die Aufgaben die durch die Visualisierung gelöst werden sollen:

**Das Problem ist folgendes:** In hochkomplexen Systemen sind die Einflussfaktoren die die Ausprägung eines Merkmals beeinflussen nicht immer klar. 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 Merkmals zu schätzen. Die klassische Herangehensweise in den Einfluss zu überprüfen sind hochkomplizierte und benötigen statistisches Hintergrundwissen auf Seiten des Anwenders. Die Visualisierungsan-

wendung soll es dem Anwender ermöglichen den Datensatz und die Merkmale derer zu untersuchen und Rückschlüsse auf die Beziehungen von Ausprägungen untereinander und mit Verhaltensindikatoren und zu treffen. Dies ermöglicht dem Anwender die Daten auf Validität zu überprüfen und die

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** 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 Zielproblems ergeben.

**Anforderungen an das Design.**

### 3.3 Präsentation der Visualisierungen

Präsentieren sie die visuelle Abbildungen und Kodierungen der Daten und Interaktionsmöglichkeiten. Sie müssen begründen, warum und wie gut ihre Designentscheidungen die erstellten Anforderungen erfüllen. Weiterhin müssen sie begründen, warum die gewählte visuelle Kodierung der Daten für das zulösenden 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. Machen sie hier nicht den Fehler, einfach nur Visualisierung aus den vorgegebenen Bereichen zu diskutieren, weil das in der Regel nicht sinnvoll ist. Wenn sie sich für einen Scatterplot entschieden haben, ist ein Zeitreihendiagramm in der Regel keine Alternative. Diskutieren sie also nicht einfach Zeitreihendiagramme, weil sie in den Anforderungen an das Projekt neben Scatterplots stehen, sondern suchen sie nach echten alternativen Visualisierungen, die zum Aufbau eines vergleichbaren mentalen Modells führen. 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.

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

#### 3.3.2 Visualisierung Zwei: Boxplott

Ziel der Visualisierung: Erklärung, wie diese aufgebaut sind. Daher

### **3.3.3 Visualisierung Drei: Force Graph.**

Ziel der Visualisierung: Erklärung, wie diese aufgebaut sind.

## **3.4 Interaktion**

Die präsentierten Visualisierungstechniken müssen interaktiv zu einer Anwendung verknüpft werden. 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. Stellen die Gliederung ihres Quellcodes vor. 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.

## **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

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

### Literatur

- [1] Tamara Munzner. “Process and Pitfalls in Writing Information Visualization Research Papers”. In: *Information Visualization: Human-Centered Issues and Perspectives*. Hrsg. von Andreas Kerren u. a. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, S. 134–153. ISBN: 978-3-540-70956-5. DOI: 10.1007/978-3-540-70956-5\_6. URL: [https://doi.org/10.1007/978-3-540-70956-5\\_6](https://doi.org/10.1007/978-3-540-70956-5_6).