

Institut für Visualisierung und Interaktive Systeme

Universität Stuttgart  
Universitätsstraße 38  
D-70569 Stuttgart

Bachelorarbeit

# **Geschichtete Diagramme oder Flächen: Eine komparative Benutzerstudie**

Julian Lang

<b>Studiengang:</b>	Softwaretechnik
<b>Prüfer/in:</b>	Prof. Dr. Thomas Ertl
<b>Betreuer/in:</b>	Moritz Knabben, M. Sc. Max Franke, M. Sc.
<b>Beginn am:</b>	13. August 2020
<b>Beendet am:</b>	13. Februar 2021



## **Kurzfassung**

In dieser Arbeit werden drei verschiedene Visualisierungstechniken auf ihre Lesbarkeit untersucht und verglichen. Betrachtet werden die Visualisierungstechniken Liniendiagramm, ThemeRiver und Small Multiples. Die Lesbarkeit wird dabei durch die benötigte Antwortzeit und die Richtigkeit gemessen. Dafür wurde eine Studie mit 48 Teilnehmer\*innen durchgeführt, deren Kern es war, drei verschiedene Aufgaben zu lösen. Es wurde untersucht, inwiefern die Teilnehmer\*innen die Aufgaben anhand von den verschiedenen Visualisierungstechniken und zwei verschiedenen Zeitreihenmengen lösen können. Die Studie ergab, dass die Wahl der Visualisierung keine große Rolle bei der Wahrnehmung der Daten spielt. Müsste man sich jedoch für eine Technik entscheiden, so eignet sich der ThemeRiver gegenüber dem Liniendiagramm und den Small Multiples besser. Der ThemeRiver wies für jede Aufgabe eine geringe Antwortzeit und zusätzlich eine hohe Korrektheit auf. Das Liniendiagramm und die Small Multiples schnitten zeittechnisch ähnlich ab, während die Korrektheit der Small Multiples in zwei von drei Aufgaben niedriger ausfiel. Nach dem ThemeRiver gab es dadurch und durch eine erhöhte Beliebtheit eine Tendenz für das Liniendiagramm.

## **Abstract**

In this thesis, three different visualization techniques are investigated and compared for their readability. The visualization techniques examined consist of the linechart, the ThemeRiver, and the small multiples. The readability is measured based on the required response time and the correctness of the answers. For this purpose, a study with 48 participants was carried out. The study consisted of three different tasks. It was investigated, how the participants can solve the tasks dependent on the different visualization techniques and two different time series counts. The study found, that the choice of visualization does not play a major role in how the data is perceived. However, if you have to decide on a technique, the ThemeRiver is better suited to the linechart and the small multiples. The ThemeRiver showed a short completion time for each task and a high level of correctness. The linechart and the small multiples performed similarly in terms of completion time, while the correctness of the small multiples was lower in two out of three tasks. After the themeriver, there was a tendency for the linechart because of this, and because of its increased popularity.



# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>vii</b>
<b>Tabellenverzeichnis</b>	<b>viii</b>
<b>Verzeichnis der Algorithmen</b>	<b>viii</b>
<b>1 Einleitung</b>	<b>1</b>
<b>2 Grundlagen</b>	<b>3</b>
2.1 (Informations-)Visualisierung . . . . .	3
2.2 Techniken zur Darstellung von Zeitreihen . . . . .	4
2.2.1 Zeitreihen . . . . .	4
2.2.2 Liniendiagramm . . . . .	4
2.2.3 Small Multiples . . . . .	6
2.2.4 Diagramm mit geschichteten Zeitreihen . . . . .	6
2.3 Wahrnehmung . . . . .	8
2.4 Randomisierung . . . . .	10
2.5 Statistische Auswertung . . . . .	10
2.5.1 Konfidenzintervalle . . . . .	10
2.5.2 Empirische Bootstrapping-Konfidenzintervalle . . . . .	12
2.5.3 Vermeidung von Nullhypothesen-Signifikanztests und p-Werten . . . . .	13
2.6 Zusammenfassung . . . . .	14
<b>3 Verwandte Arbeiten</b>	<b>15</b>
<b>4 Benutzerstudie</b>	<b>19</b>
4.1 Aufgaben . . . . .	19
4.1.1 Aufgabe 1: Vergleich von mehreren Werten (Maximum) . . . . .	19
4.1.2 Aufgabe 2: Vergleich von mehreren Flächen (Area) . . . . .	20
4.1.3 Aufgabe 3: Vergleich von Summen (Global) . . . . .	20
4.2 Erwartete Effekte . . . . .	21
4.3 Teilnehmer*innen . . . . .	22
4.4 Apparat . . . . .	23
4.5 Szenario . . . . .	23
4.6 Datensätze . . . . .	24
4.7 Studienablauf . . . . .	26
4.8 Studiendesign . . . . .	27
4.9 Experimentelle Bedingungen . . . . .	27
4.9.1 Farbschtest und Farbwahl . . . . .	28
4.9.2 Visualisierungstechniken . . . . .	29

4.9.3	Anzahl der Zeitreihen . . . . .	29
4.10	Zusammenfassung . . . . .	30
<b>5</b>	<b>Vorstudien</b>	<b>31</b>
5.1	Erste Vorstudie . . . . .	31
5.1.1	Probleme . . . . .	31
5.1.2	Sonstiges Feedback . . . . .	32
5.2	Zweite Vorstudie . . . . .	32
5.2.1	Probleme . . . . .	32
5.2.2	Sonstiges Feedback . . . . .	32
5.3	Zusammenfassung . . . . .	33
<b>6</b>	<b>Ergebnisse und Diskussion</b>	<b>35</b>
6.1	Auswertung und Diskussion der Daten . . . . .	35
6.1.1	Ebene 1 . . . . .	36
6.1.2	Ebene 2 . . . . .	37
6.1.3	Ebene 3 . . . . .	42
6.2	Feedback der Teilnehmer*innen . . . . .	45
6.3	Vergleich von Ergebnissen verwandter Arbeiten . . . . .	47
<b>7</b>	<b>Zusammenfassung</b>	<b>49</b>
<b>A</b>	<b>Anhang</b>	<b>51</b>
A.1	Ergebnisse . . . . .	52
A.2	Ablauf der Benutzerstudie . . . . .	53
A.3	Demographische Daten . . . . .	61
A.4	Sonstiges . . . . .	64
A.4.1	Auswertungen . . . . .	64
A.4.2	Farbpalette . . . . .	65
	<b>Literaturverzeichnis</b>	<b>67</b>

# Abbildungsverzeichnis

2.1	Liniendiagramm mit vier Zeitreihen. . . . .	5
2.2	Darstellung von höherer Unordnung. . . . .	5
2.3	Darstellungen von Small Multiples. . . . .	6
2.4	Wahrnehmung der Distanz bei Flächen . . . . .	7
2.5	Stacked Area Chart mit vier Zeitreihen. . . . .	7
2.6	ThemeRiver mit vier Zeitreihen. . . . .	8
2.7	<i>Elementary perceptual tasks</i> nach Cleveland et al. . . . .	9
2.8	Darstellung von Konfidenzintervallen. . . . .	11
4.1	Darstellungen der Aufgabe Maximum. . . . .	20
4.2	Darstellungen der Aufgabe Area. . . . .	20
4.3	Darstellungen der Aufgabe Global. . . . .	21
4.4	Darstellung der Ishihara-Tafel. . . . .	28
6.1	Darstellung der Ebenen für die Auswertung. . . . .	36
6.2	Auswertung der ersten Ebene. . . . .	37
6.3	Auswertung der zweiten Ebene. . . . .	38
6.4	Ergebnisse der Teilnehmer*innen mit und ohne Rot-Grün-Farbsehschwäche. . . . .	40
6.5	Alternative Darstellung d. Zeiten (mit und ohne Farbsehschwäche). . . . .	41
6.6	Alternative Darstellung d. Korrektheit (mit und ohne Farbsehschwäche). . . . .	41
6.7	Auswertung der dritten Ebene – gruppiert nach Aufgaben. . . . .	43
6.8	Auswertung der dritten Ebene – gruppiert nach Zeitreihenanzahl. . . . .	44
6.9	(Un-)Beliebtheit der Visuslisierungstechniken. . . . .	46
6.10	Alt. Darstellung der (Un-)Beliebtheit der Visuslisierungstechniken. . . . .	47
A.1	Studie: Einführungsbildschirm. . . . .	53
A.2	Studie: Demographischer Fragebogen. . . . .	54
A.3	Studie: Skalierungsaufgabe. . . . .	54
A.4	Studie: Ishihara-Tafel. . . . .	55
A.5	Studie: Beschreibung der ersten Aufgabe. . . . .	56
A.6	Studie: Übungsdurchlauf der ersten Aufgabe. . . . .	56
A.7	Studie: Beschreibung der ersten Aufgabe (2). . . . .	57
A.8	Studie: Liniendiagramm bei erster Aufgabe. . . . .	57
A.9	Studie: Beschreibung der zweiten Aufgabe. . . . .	58
A.10	Studie: Beispiele der zweiten Aufgabe. . . . .	58
A.11	Studie: Beschreibung der dritten Aufgabe. . . . .	59
A.12	Studie: Beispiele der dritten Aufgabe. . . . .	59
A.13	Studie: Feedback-Fragebogen. . . . .	60
A.14	Studie: Abschlussbildschirm. . . . .	60
A.15	Wordcloud der Berufe der Teilnehmer*innen. . . . .	61

A.16 Altersverteilung der Teilnehmer*innen. . . . .	61
A.17 Verteilung des höchsten Abschlusses der Teilnehmer*innen. . . . .	62
A.18 Verteilung der Wahrnehmungseinflüssen . . . . .	63
A.19 Auswertung der dritten Ebene mit vier Zeitreihen. . . . .	64
A.20 Auswertung der dritten Ebene mit sieben Zeitreihen. . . . .	64
A.21 Farbpalette für vier und sieben Zeitreihen. . . . .	65

## Tabellenverzeichnis

A.1 Tabelle mit den Ergebnissen der ersten Ebene. . . . .	52
A.2 Tabelle mit den Ergebnissen der zweiten Ebene. . . . .	52
A.3 Tabelle mit den Ergebnissen der dritten Ebene mit vier Zeitreihen. . . . .	52
A.4 Tabelle mit den Ergebnissen der dritten Ebene mit sieben Zeitreihen . . . . .	52

## Verzeichnis der Algorithmen

4.1 Generierung der Random Walk-Schritte. . . . .	25
4.2 Normierung der Random Walk-Schritte. . . . .	25

# 1 Einleitung

Das Schichten von Flächen in einem Diagramm ist eine beliebte Art der Visualisierung von Daten in Form von mehreren Zeitreihen. Neben dem Schichten von Flächen gibt es weitere bekannte und beliebte Visualisierungstechniken wie beispielsweise das Liniendiagramm oder das Balkendiagramm. Diese Arten von Diagrammen gehören mittlerweile zum Alltag des Menschen und tauchen in vielen verschiedenen Bereichen auf. Beispiele findet man im Finanzwesen (z.B. Einnahmen, Aktienkurse), in der Medizin (z.B. Geburtenrate, Sterberate), in der Politik (z.B. Einwohnerzahl, Wahlergebnisse) oder im Sport (z.B. Tabellenplatzierung über eine Saison hinweg, gefallene Tore).

Hierbei stellen die Visualisierungstechniken die Zeitreihen in einem gemeinsamen Diagramm dar. Alternativ können solche Zeitreihen auch in einzelnen Diagrammen dargestellt werden. Das bedeutet, dass jede Zeitreihe ein eigenes Diagramm verwendet. Welche Visualisierungstechnik sich am besten eignet, hängt dabei von den zu Grunde liegenden Daten ab und davon, welche Information(en) die Betrachter\*innen aus den gezeigten Daten extrahieren sollen. Jede Visualisierungstechnik bringt dabei Vor- und Nachteile mit sich. Eine Visualisierungstechnik eignet sich dementsprechend gut, wenn die Betrachter\*innen die zu absolvierende Aufgabe möglichst schnell und korrekt lösen können.

Das Ziel dieser Arbeit besteht darin, verschiedene Visualisierungstechniken zu untersuchen und zu vergleichen. Dabei wird untersucht, wie sich die Lesbarkeit in Abhängigkeit der Anzahl der Zeitreihen und der zu absolvierenden Aufgabe verändert. Die Lesbarkeit einer Visualisierung kann man als Maß dafür verstehen, inwiefern eine Visualisierung die Betrachter\*innen bei der Wahrnehmung von Daten unterstützt. Um die verschiedenen Visualisierungstechniken zu untersuchen und zu vergleichen, wurde eine Studie mit 48 Teilnehmer\*innen von der Crowdsourcing-Plattform Prolific durchgeführt. Die Untersuchung der Lesbarkeit erfolgte anhand von drei Visualisierungstechniken. Dabei mussten die Teilnehmer\*innen drei verschiedene Aufgaben mit den jeweiligen Visualisierungstechniken lösen. Die aufgezeichneten Antwortzeiten, die Korrektheit und die demographischen Daten waren Grundlage für die anschließende Auswertung. Die Zeiten wurden mit Hilfe von statistischen Methoden ausgewertet und in Form von Konfidenzintervallen dargestellt, während die Korrektheit durch die relative Häufigkeit der korrekten Antworten definiert ist. Anschließend wurden die Ergebnisse miteinander verglichen. Zusätzlich wurde untersucht, wie sich die Farbwahrnehmung der Personen mit und ohne Farbsehschwäche auf die Ergebnisse der Diagramme und Aufgaben auswirkte.



## 2 Grundlagen

Dieses Kapitel umfasst Themengebiete und Definitionen von Begriffen, die für das Verständnis dieser Arbeit von Bedeutung sind. Im Folgenden wird beschrieben, was man unter einer Visualisierung versteht. Dabei wird insbesondere auf die Visualisierung von Informationen eingegangen. Daran schließt sich die Erklärung der für diese Arbeit wichtigsten Techniken für die Darstellung von mehreren Zeitreihen an. Zu diesen Techniken gehört oftmals der Gebrauch von Farben. Da Farben nicht immer gleich wahrgenommen werden und es Personen gibt, die unter einer Farbsehschwäche leiden, wird ein kurzer Überblick über die häufigsten Farbsehschwächen dargelegt. Der letzte Abschnitt dieses Kapitels beschreibt die statistischen Methoden, die für die Auswertung der Ergebnisse angewandt wurden.<sup>1</sup>

### 2.1 (Informations-)Visualisierung

Für den Begriff der (*Informations-*)*Visualisierung* gibt es keine einheitliche, allgemeingültige Definition. Keim et al. [KMSZ06] definieren den Begriff der Informationsvisualisierung beispielsweise folgendermaßen:

“We define information visualization more generally as the communication of abstract data relevant in terms of action through the use of interactive visual interfaces.” [KMSZ06]

Nach Däßler [Däß99] werden „unter Informationsvisualisierung [...] heute alle Konzepte, Methoden und Tools zur visuellen Darstellung von Informationen aus Datenbanken, digitalen Bibliotheken oder anderen großen Dokumentsammlungen, zusammengefaßt.“

Man kann unter einer Visualisierung also den Prozess zur Darstellung von Informationen, Daten und/oder Prozessen verstehen. Zur Darstellung solcher Informationen gehört der Einsatz von statischen, visuellen Grundkomponenten wie beispielsweise Linien, Punkten, Balken und Diagrammen [BVB+13]. Mit Hilfe dieser Komponenten lassen sich graphische Gesamtbilder erzeugen, aus denen neue Erkenntnisse gewonnen werden können [FvWSN08]. Besonderheiten wie Extremwerte, Tendenzen und Muster, die sonst aus den rohen Daten schwerer auszulesen wären, lassen sich dadurch besser erkennen. Im Gebiet der Informationsvisualisierung läuft ein solcher Prozess zumeist mit Hilfe digitaler Geräte ab [KMSZ06].

Neben den statischen visuellen Elementen gibt es viele zusätzliche Möglichkeiten zur Darstellung von Daten. Darunter fallen Aspekte wie die Nutzung von Interaktion, Farben oder dynamischen Elementen (z.B. Animation) [Kos07; Man10]. Das Hinzuziehen solcher Möglichkeiten hängt dabei immer von den darzustellenden Daten ab.

---

<sup>1</sup>Diverse Begriffe wurden zur Verdeutlichung kursiv geschrieben.

Der Prozess der Visualisierung beginnt mit der Aufbereitung der rohen Daten und der anschließenden Umwandlung in graphische Objekte. Dabei wird auch die Person involviert, die eine solche graphische Darstellung von Daten anschließend betrachtet. Die Perzeption und die daraus resultierende Extraktion der dargestellten Daten kann je nach Person variieren. Dazu tragen Faktoren wie Vorwissen und Perzeptionsfähigkeiten [vWij06], aber auch biologische Aspekte in Form von z.B. Sehschwächen bei. Für die (Informations-) Visualisierung ist dies von großer Bedeutung, da der ideale Fall darin besteht, dass die Extraktion der Daten ein objektiver Prozess ist. Das bedeutet, dass jede Person dieselben Informationen aus den Visualisierungen extrahieren soll. Durch die genannten Faktoren ist das Erreichen dieser Objektivität jedoch nur schwer möglich.

## 2.2 Techniken zur Darstellung von Zeitreihen

Für die Darstellung von Zeitreihen existieren verschiedene Techniken. Darunter fallen unter anderem Liniendiagramme, Balkendiagramme und geschichtete Diagramme. Da Menschen, wie anfangs angesprochen, Daten subjektiv aus den Diagrammen extrahieren, existiert für die Darstellung von Daten keine perfekte Technik. Trotz der subjektiven Empfindung einer Visualisierung existieren auch objektive Eigenschaften wie beispielsweise die Diagrammgröße oder die Anzahl der Zeitreihen, die sich als Vor- und Nachteile herausstellen und abgewägt werden müssen. In den folgenden Abschnitten wird sowohl kurz erläutert, was man unter einer Zeitreihe versteht, als auch die für diese Arbeit wichtigen Darstellungstechniken beschrieben. Dabei ist anzumerken, dass dabei ausschließlich die zweidimensionale Darstellung der Diagramme betrachtet wird, da diese Darstellung im Bereich der Informationsvisualisierung üblich ist.

### 2.2.1 Zeitreihen

Eine *Zeitreihe* beschreibt eine Menge von Datenpunkten in einer zeitlichen Anordnung. Eine Menge an Datenpunkten ist zum Beispiel die Einwohnerzahl eines Landes. Dabei ist jedem zeitlichen Wert mindestens ein Wert der Datenmenge zugewiesen. Die Einwohnerzahl über einen gewissen Zeitraum (z.B. Tage, Wochen oder Jahre) hinweg wird als Zeitreihe beschrieben. Solche Zeitreihen finden beispielsweise Anwendungen im Finanzwesen [Tsa05] und in der Medizin [CA15; LKAV05]. Die Untersuchung solcher Zeitreihen wird als *Zeitreihenanalyse* bezeichnet.

### 2.2.2 Liniendiagramm

Das *Liniendiagramm* ist eine der grundlegendsten Techniken zur Darstellung von multivariaten temporalen Daten. Diese Technik reicht zurück bis in das späte 18. Jahrhundert, in dem sie von William Playfair erfunden wurde [Spe06; SW06].

In einem Liniendiagramm kann eine beliebige Anzahl von Zeitreihen dargestellt werden. Die Darstellung solcher Zeitreihen erfolgt dadurch, dass man die Zeitdaten auf der X-Achse und die dazugehörigen Werte auf der Y-Achse abbildet. Die Wertepaare aus X- und Y-Werten

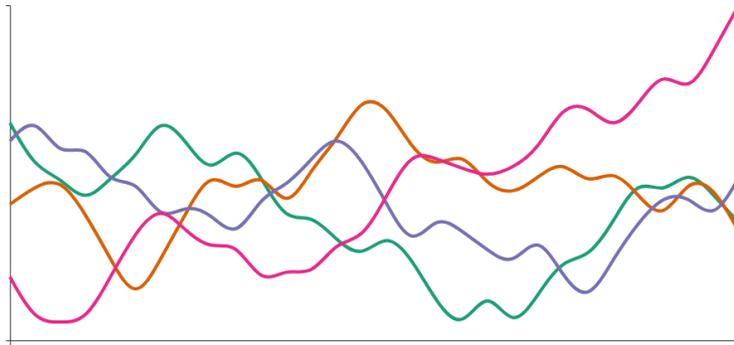


Abbildung 2.1: Darstellung eines Liniendiagramms mit vier Zeitreihen.

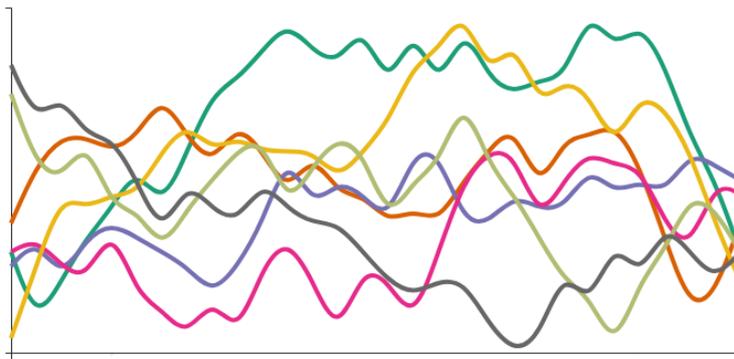


Abbildung 2.2: Beispielhafte Darstellung von höherer Unordnung anhand eines Liniendiagramms mit sieben Zeitreihen.

ergeben eine Anzahl von Punkten, die in dem Diagramm mit Hilfe einer Linie verbunden werden. Zur besseren Unterscheidung zwischen den verschiedenen Zeitreihen wird den Linien eine gewisse graphische Eigenschaft wie zum Beispiel eine eigene Farbe zugewiesen. Bei einem Liniendiagramm handelt es sich um eine *Shared Space*-Technik [JME10], da hierbei alle Zeitreihen in einem einzelnen Diagramm dargestellt werden.

In Abbildung 2.1 ist ein Liniendiagramm mit vier verschiedenfarbigen Zeitreihen zu sehen. Die Darstellung mehrerer Zeitreihen in einem Liniendiagramm ermöglicht direkte Vergleiche zwischen den Zeitreihen. Ebenso lassen sich Besonderheiten wie extreme Werte(-bereiche) leicht ausfindig machen. Bei einer hohen Anzahl an Zeitreihen kann das Diagramm jedoch durch die Überschneidungen der Linien unübersichtlich werden. Dies wird auch *Visual Clutter* genannt. Eine beispielhafte Darstellung ist in Abbildung 2.2 zu sehen.

Neben der statischen Darstellung von Zeitreihen in einem Liniendiagramm gibt es ebenso dynamische Darstellungsmöglichkeiten. Darunter fällt der Gebrauch von Animationen und/oder Interaktion. Um ein mögliches Beispiel zu liefern, könnte man mit der Maus über einen gewissen Punkt einer Zeitreihe fahren und dabei den Wert dieses Zeitpunkts in einem kleinen Fenster angezeigt bekommen.

### 2.2.3 Small Multiples

*Small Multiples* sind eine allgemeine Technik, bei der mehrere Grafiken mit gleichen räumlichen Ausmaßen unterschiedliche Daten zeigen. Hierbei werden die einzelnen Zeitreihen nicht in einem einzigen Diagramm, sondern in mehreren Diagrammen dargestellt. Diese Technik nennt man *Split Space*-Technik. Die einzelnen Diagramme können untereinander, nebeneinander oder als Gitter angeordnet werden und besitzen die gleichen Maße [HKA09]. Da nun keine Überschneidungen von Linien mehr vorliegen, gibt es die Möglichkeit, die unterliegende Fläche der Linien in der dazugehörigen Farbe einzufärben, um die Betrachtung der einzelnen Diagramme zu erleichtern [JME10]. Die einheitliche Größe der Diagramme ermöglicht das Vergleichen der Zeitreihen. Dabei ist es jedoch wichtig, dass die Achsen der Diagramme allesamt einheitlich sind.

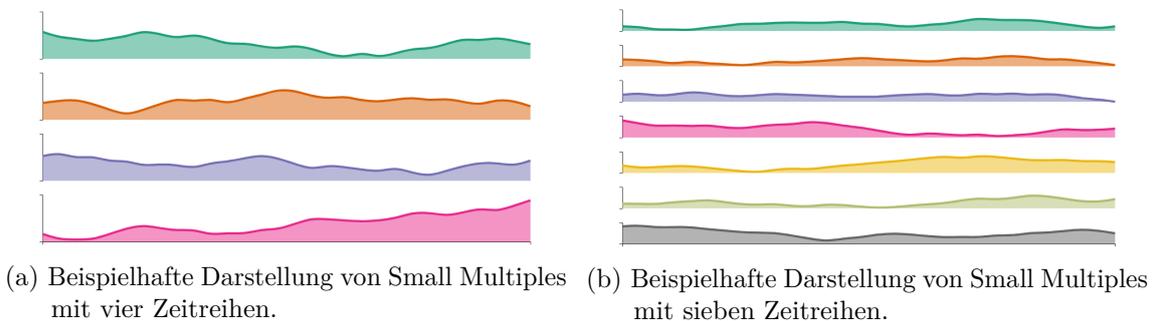


Abbildung 2.3: Zwei beispielhafte Darstellungen von Small Multiples mit verschiedener Anzahl an Zeitreihen.

Zwei Darstellungen von Small Multiples sind in Abbildung 2.3 zu sehen. Bei der Darstellung der Small Multiples ist zu beachten, dass der vertikale Gesamtplatz unter den einzelnen Diagrammen aufgeteilt wird. Bei einer hohen Anzahl an Zeitreihen verringert sich somit der Platz für jedes einzelne Diagramm, wodurch Vergleiche wiederum erschwert werden können. Der Unterschied des Platzes ist in den beiden Abbildungen 2.3a und 2.3b zu sehen.

### 2.2.4 Diagramm mit geschichteten Zeitreihen

Bei diesen Techniken werden einzelne Zeitreihen übereinandergestapelt, die sich den Platz eines Diagramms somit teilen. Dadurch erhält man mehrere übereinanderliegende Flächen, die meist mit einer Farbe versehen werden. Hierbei gibt es mehrere Möglichkeiten, diese geschichteten Flächen darzustellen. Byron und Wattenberg [BW08] nennen mehrere Faktoren, die sich auf die Lesbarkeit auswirken so wie beispielsweise die Basislinie und der *Wiggle*-Faktor, also die Schwingung der Flächen. Neben der Lesbarkeit spielt bei dieser Technik auch die Ästhetik eine wichtige Rolle. Details zu diversen Techniken sind in den folgenden Unterkapiteln erläutert.

Diese Techniken eignen sich für Daten, bei denen es sich als sinnvoll erweist, sie gegebenenfalls zu aggregieren und zu vergleichen (z.B. Einnahmen von Verkäufen). Jedoch gibt es auch Datentypen wie Temperaturen, für die sich die Aggregation und somit diese Technik nicht eignet [HKA09]. Nach Cleveland und McGill [CM84] umfasst das Vergleichen von

Werten eher die Größenvergleiche, also die visuelle Höhe einer Zeitreihe, als die genauen Positionsvergleiche. Ein Problem dieser Visualisierungstechnik ist, dass sich Betrachter\*innen oftmals fehlleiten lassen, da sie eher die orthogonale Distanz zwischen der unteren und oberen Grenze der Flächen wahrnehmen als die vertikale Distanz [BZW+21; CM84]. Diese mögliche Wahrnehmung der Zeitreihen ist in Abbildung 2.4 dargestellt.

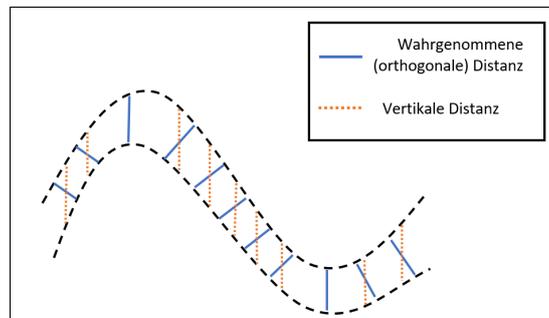


Abbildung 2.4: Die Perzeption führt eher zur Wahrnehmung der orthogonalen Distanz anstatt der vertikalen Distanz - Abbildung angelehnt an Bu et al. [BZW+21].

### Stacked Area Chart

Beim *Stacked Area Chart* verwendet jede Zeitreihe die Werte der unterliegenden Zeitreihe als Basislinie. Da unter der untersten Zeitreihe keine Zeitreihe liegt, verwendet diese die X-Achse als Basislinie und besitzt dadurch als einzige Zeitreihe keine verzerrte Basislinie. Dadurch entstehen mehrere übereinandergeschichtete Flächen, die zur besseren Unterscheidung in verschiedenen Farben eingefärbt werden [HKA09; JME10]. Die Nutzung dieser Art von Basislinien ist nur möglich, da durch die Stapelung der Zeitreihen keine Überschneidungen möglich sind.

In Abbildung 2.5 ist ein Stacked Area Chart mit vier geschichteten Zeitreihen zu sehen. Die Höhe einer einzelnen Fläche zu einem beliebigen X-Wert beschreibt den Wert der Zeitreihe zu diesem Zeitpunkt.



Abbildung 2.5: Beispielhafte Darstellung eines Stacked Area Charts mit vier Zeitreihen.

## Streamgraph/ThemeRiver

Der *ThemeRiver* ist eine spezielle Art des Stacked Area Charts. Bei einem ThemeRiver werden die einzelnen Zeitreihen ebenfalls übereinandergeschichtet, jedoch liegen diese im Gegensatz zum Stacked Area Chart dabei mittig um eine zentrale Achse [TWP+16]. Das bedeutet, dass diese aufwärts und abwärts der Achse gestapelt werden, anstatt wie im Stacked Area Chart nur aufwärts einer Achse. Durch die Anordnung um die zentrale Achse erhält der gesamte Graph eine fließende, symmetrische Form. Zur besseren und eindeutigen Erkennung der Zeitreihen besitzt dabei jede Zeitreihe eine visuelle Eigenschaft in Form einer eigenen Farbe. Wichtig ist auch, dass sich mit einem ThemeRiver ausschließlich positive Werte darstellen lassen. Ein Beispiel für einen solchen ThemeRiver ist in Abbildung 2.6 zu sehen.

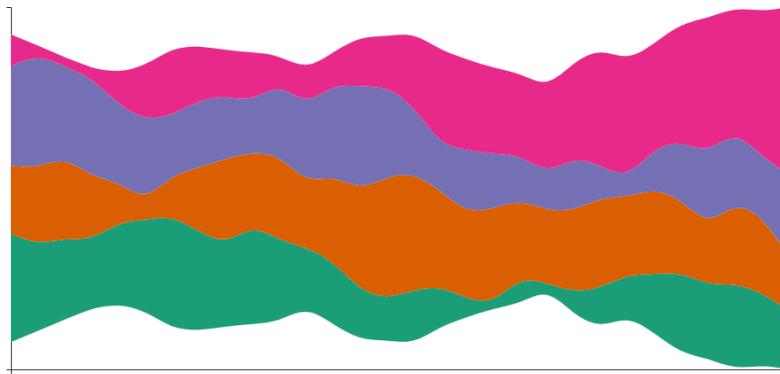


Abbildung 2.6: Beispielhafte Darstellung eines ThemeRivers mit vier Zeitreihen.

Mit Hilfe des ThemeRivers können sich die Betrachter\*innen einen Überblick über die Veränderungen der Zeitreihen machen. Für das Erkennen von kleineren Trends ist der ThemeRiver jedoch weniger geeignet. Der Nachteil des ThemeRivers besteht darin, dass dieser die mittig gelegenen Zeitreihen stärker hervorhebt als die außen gelegenen.

Verwandt mit dem ThemeRiver ist der *Streamgraph*. Dieser ordnet die einzelnen Zeitreihen nicht um eine mittige Achse an, sondern mit dem Ziel, die Verzerrung der Zeitreihen zu verringern. Um dies zu erreichen, wird der Wiggle-Faktor reduziert [BW08; BZW+21]. Der gesamte Graph wird dadurch jedoch asymmetrisch. Eines der bekanntesten Beispiele für eine Streamgraphvisualisierung ist die Darstellung von Filmeinkommen im Artikel „The Ebb and Flow of Movies: Box Office Receipts 1986 — 2008“ der New York Times von Mathew Bloch, Lee Byron, Shan Carter und Amanda Cox [BBCC08].

## 2.3 Wahrnehmung

Für die Wahrnehmung und Interpretation von Visualisierungen spielt die menschliche visuelle Wahrnehmung einen wichtigen Faktor. Darunter versteht man die Extraktion oder das visuelle Decodieren von Informationen aus Graphen und anderen Darstellungen. Cleveland et al. sprechen in ihrer Arbeit [CM84] von „elementary perceptual tasks“, die bei der Extraktion von quantitativen Informationen aus Graphen genutzt werden. Darunter

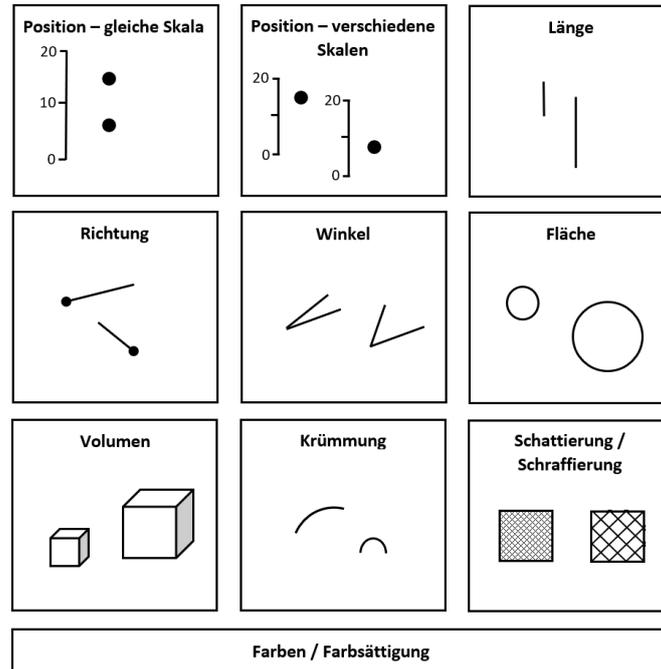


Abbildung 2.7: *Elementary perceptual tasks*, die für die Extraktion von Informationen aus Graphen oder anderen Darstellungen genutzt werden – Reproduzierte Abbildung nach Cleveland et al. [CM84].

versteht man die in Abbildung 2.7 dargestellten Eigenschaften, von denen die Betrachter\*innen bei der Extraktion von Informationen, bzw. Daten, mindestens eine nutzen. Beispielsweise nutzen die Betrachter\*innen bei einem Balkendiagramm die Position, die Länge und möglicherweise auch die Fläche eines Balkens, um Informationen herauszufiltern. Für die Extraktion von Informationen aus den in dieser Arbeit vorkommenden Visualisierungen wurden je nach Aufgabe (siehe Kapitel 4.1) verschiedene Eigenschaften aus Abbildung 2.7 benötigt.

Neben Größe, Länge, Position und weiteren Eigenschaften spielen auch Farben eine Rolle (siehe Abbildung 2.7). Diese verkörpern eine wichtige Eigenschaft sowohl zur Extraktion von Informationen (Farbsaturierungen, beispielsweise in Form von Temperaturen:  $0^{\circ}\text{C}$  = blau bis  $50^{\circ}\text{C}$  = rot), als auch zur Unterscheidung von Graphen. Da Farben, im Gegensatz zu Längen oder Positionen, jedoch subjektiv und somit von jeder Person anders wahrgenommen werden, ist es sehr schwierig, eine geeignete Farbgebung für Graphen zu finden. Hinzu kommen diverse, nicht beeinflussbare Faktoren wie Farbsehschwächen – insbesondere die Rot-Grün-Farbsehschwäche.

Statistisch ist die Rot-Grün-Sehschwäche die am häufigsten auftretende Farbsehschwäche. Von dieser sind etwa 8% aller Männer und 0,5% aller Frauen nordwestlicher europäischer Abstammung betroffen [Nat15; Won11]. Ihr folgt die Blau-Gelb-Sehschwäche und die vollständige Farbenblindheit. In dieser Arbeit sind Farben zur Unterscheidung von Zeitreihen von hoher Wichtigkeit. Daher wurde ein hoher Wert auf eine möglichst geeignete Farbpalette gelegt. Weitere Vorgänge, die die Wahl der Farbpalette betreffen, sind in den Kapiteln 4.9.1 und 5 beschrieben.

## 2.4 Randomisierung

*Randomisierung* findet heutzutage überwiegend in klinischen Studien Anwendung und zählt dort mittlerweile auch als standardisierte Methode der empirischen Erkenntnisgewinnung [LI19]. Anhand der Randomisierung werden Versuchspersonen zufällig in verschiedene Versuchsgruppen eingeteilt. Dadurch soll das Auftreten eines möglichen *Selection Bias* so gut wie möglich verhindert werden. Zudem fördert die Randomisierung die Vergleichbarkeit der Versuchsgruppen, da alle bekannten oder unbekannt Faktoren, die Einfluss auf das Ergebnis haben können, auf beinahe dieselbe Weise auf die Gruppen aufgeteilt werden. Anhand dessen ist es überhaupt erst möglich, Vergleiche zu tätigen und Erkenntnisse zu gewinnen.

Für die Randomisierung existieren mehrere mögliche Methoden wie beispielsweise die *einfache Randomisierung*, die *Block-Randomisierung* oder die *adaptive Randomisierung* [LI19]. Bei ersterem werden Versuchsteilnehmer\*innen beispielsweise mittels eines Münz- oder Würfelwurfs einer Gruppe zugeordnet. Der Vorteil ist hierbei, dass jegliche Bias durch die Beseitigung der Vorhersagbarkeit minimiert werden. Jedoch ist es möglich, dass dadurch ein Ungleichgewicht bei der Anzahl der Personen in den Gruppen entsteht. Somit würde die Aussagekraft der Studie sinken. Eine gewisse Balancierung der Teilnehmerzahlen wird als plausibel erachtet, jedoch gestaltet sich diese nicht immer einfach.

In dieser Arbeit soll die Randomisierung dafür sorgen, dass für alle Teilnehmer\*innen die gleichen Versuchsbedingungen gelten, während ein Bias weitestgehend vermieden werden soll. Anhand dessen wurde für die Reihenfolge der Aufgaben, die Wahl der Visualisierung und die Wahl der Zeitreihenanzahl jeweils eine einfache Randomisierung vorgenommen. Um ein mögliches Ungleichgewicht zwischen den jeweiligen Verhältnissen zu vermeiden, wurde darauf geachtet, dass ein festes Verhältnis zwischen den Visualisierungsarten (1:1:1) und der Zeitreihenanzahl (1:1) eingehalten wurde.

## 2.5 Statistische Auswertung

Um die durch die Studie erhaltenen Daten auszuwerten, wurden bestimmte Auswertungsmethoden verwendet. Die genutzten Methoden werden in den folgenden Unterkapiteln beschrieben. Diese sind für das Verständnis von Kapitel 6 notwendig. Zudem wird erläutert, wieso von der Nutzung von Signifikanztests mit Nullhypothesen und  $p$ -Werten abgesehen wurde.

### 2.5.1 Konfidenzintervalle

Ein *Konfidenzintervall* beschreibt ein statistisches Intervall, mit dem der wahre Parameter  $\mu$ , also der wahre Mittelwert, in der Grundgesamtheit mit einer bestimmten Wahrscheinlichkeit gefunden werden soll. Dieses Intervall basiert auf einer Stichprobe, aus der der Parameter  $\hat{\mu}$  gezogen wird. Dieser Parameter beschreibt dabei lediglich den Mittelwert der Stichprobe, um den das Intervall liegt.

Das Grundprinzip besteht darin, eine beliebige Anzahl von Stichproben zu ziehen und daraus die dazugehörigen Konfidenzintervalle aufzustellen. Um die statistische Relevanz zu erhalten, wird von den bekannten Parametern jeder Stichprobe auf den wahren Parameter  $\mu$  in der Grundgesamtheit geschlossen. Der wahre Parameter  $\mu$  kann jedoch nie exakt bestimmt werden, sondern muss geschätzt werden [Dra16]. Daher wird ein Bereich – das Konfidenzintervall – bestimmt, in dem dieser Parameter zu einer gewissen Wahrscheinlichkeit liegt. Diese Wahrscheinlichkeit wird *Konfidenzniveau* genannt und beschreibt, mit welcher Wahrscheinlichkeit der wahre Parameter in dem Konfidenzintervall liegt. Das Konfidenzniveau wird auch mit  $(1 - \alpha)$  beschrieben. Dabei beschreibt  $\alpha$  die Wahrscheinlichkeit, dass der wahre Parameter nicht in dem Intervall liegt. Typische Werte für das Konfidenzniveau sind 90%, 95% und 99%.

Ein 95% Konfidenzintervall ist also ein Intervall, in dem der wahre Parameter  $\mu$  mit einer Wahrscheinlichkeit von 95% liegt – vorausgesetzt, dieses Intervall wird mit neuen Stichproben mehrmals neu berechnet.

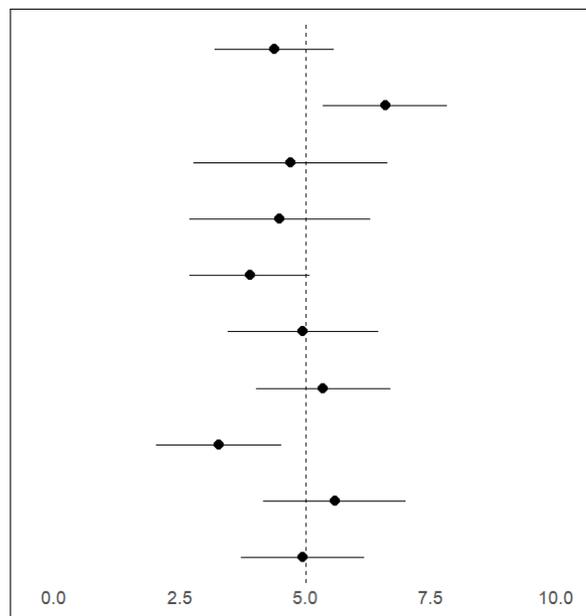


Abbildung 2.8: Beispielhafte Darstellung von zehn Konfidenzintervallen und dem wahren Mittelwert der Grundgesamtheit, basierend auf dem R-Code von Dragicevic [Dra16]. Der wahre Mittelwert ist hierbei mit  $\mu = 5$  bekannt. In acht der zehn Konfidenzintervalle ist der wahre Mittelwert enthalten, in zwei davon jedoch nicht.

Abbildung 2.8 zeigt eine beispielhafte Darstellung von zehn Konfidenzintervallen aus zehn Stichproben. Die schwarzen Punkte innerhalb der Intervalle werden *Punktschätzer* genannt und beschreiben den Mittelwert  $\hat{\mu}$  jeder Stichprobe. In der Abbildung wird angenommen, dass der wahre Mittelwert  $\mu$  bekannt ist ( $\mu = 5$ ). Anhand dessen ist ersichtlich, dass in acht der zehn Konfidenzintervalle der wahre Mittelwert in dem Intervall liegt, in zwei davon jedoch nicht.

Um ein Konfidenzintervall bestimmen zu können, müssen sowohl der Mittelwert  $\hat{\mu}$  und die Stichprobengröße  $n$ , als auch entweder die Stichprobenvarianz  $\hat{\sigma}^2$  oder die Stichprobenstandardabweichung  $\hat{s}$  gegeben sein. Da jedes Konfidenzintervall eine untere und eine obere Grenze besitzt, müssen diese mit Hilfe des sogenannten *z-score* bestimmt werden.

Der *z-score* beschreibt die Abweichung eines Rohwertes vom Mittelwert und gilt als standardisierte Zufallsvariable. Je nach Konfidenzniveau fällt die Variable  $z$  dabei anders aus. Für ein 95% Konfidenzniveau beträgt der *z-score* beispielsweise 1,96.

Ein Konfidenzintervall lässt sich mit den vorhandenen Variablen nun folgendermaßen berechnen:

$$\left[ \hat{\mu} - z \cdot \sqrt{\frac{\hat{\sigma}^2}{n}}, \hat{\mu} + z \cdot \sqrt{\frac{\hat{\sigma}^2}{n}} \right]$$

Für eine beispielhafte Berechnung eines 95% Konfidenzintervalls werden folgende Werte angenommen:

$$\begin{aligned} \hat{\mu} &= 91, & z &= 1,96, \\ \hat{\sigma}^2 &= 23,04, & n &= 67. \end{aligned}$$

Berechnung der unteren Grenze:

$$91 - 1,96 \cdot \sqrt{\frac{23,04}{67}} \approx 89,85.$$

Berechnung der oberen Grenze:

$$91 + 1,96 \cdot \sqrt{\frac{23,04}{67}} \approx 92,15.$$

Daraus ergibt sich, dass der wahre Mittelwert  $\mu$  der Grundgesamtheit zu 95% zwischen 89,85 und 92,15 liegen könnte.

## 2.5.2 Empirische Bootstrapping-Konfidenzintervalle

Das *empirische Bootstrapping* beschreibt eine statistische Methode, bei der mit Hilfe eines Computers viele verschiedene alternative Stichproben anhand der ursprünglichen Stichprobe generiert werden [Dra16]. Wie auch bei der klassischen Berechnung besteht das Ziel des Bootstrappings darin, Konfidenzintervalle zu schätzen [OB14].

Dragicevic [Dra16] beschreibt diese Art der Konfidenzintervallberechnung als neuartig, da diese einen Computer benötigt. Jedoch ist diese Methode sehr vielseitig und auf viele Verteilungen anwendbar. Dadurch ist die Notwendigkeit einer vorliegenden Normalverteilung nicht vorhanden.

Bei der Generierung der alternativen Stichproben werden aus der ursprünglichen Stichprobe viele Stichproben mit Zurücklegen gezogen. Das bedeutet, dass eine Beobachtung öfter in einer alternativen Stichprobe enthalten sein kann. Die alternativen Stichproben bestehen

dabei aus der gleichen Anzahl von Beobachtungen wie die ursprüngliche Stichprobe. Da die Anzahl der alternativen Stichproben oftmals bei über 1.000 liegt, wird für diese Methode ein Computer benötigt [OB14].

Auf Basis der vielen alternativen Stichproben erfolgt die Analyse der Daten. Hierbei liegt nun genau ein Stichprobenkennwert in Form eines Mittelwertes pro alternative Stichprobe vor. Aus der Menge an Stichprobenmittelwerten werden anhand  $\alpha$  die  $\frac{\alpha}{2}$  kleinsten und größten Mittelwerte bestimmt. Für ein Konfidenzniveau von 95% werden die 2,5% kleinsten und größten (2,5% und 97,5% Perzentil) Mittelwerte bestimmt. Diese bilden somit die Unter- und Obergrenze der zu beachtenden Mittelwerte und somit auch die Grenzen des Konfidenzintervalls.

### 2.5.3 Vermeidung von Nullhypothesen-Signifikanztests und $p$ -Werten

Ein *Nullhypothesensignifikanztest* ist eine Methode zur statistischen Erkenntnisgewinnung, bei der eine Beobachtung gegen eine Nullhypothese getestet wird [Per17]. Diese behauptet dabei, dass es bei der Beobachtung keinen Effekt gibt. Nach Dragicevic [Dra16] spricht man dabei auch oft von der „hypothesis of zero effect“. Die Methode ermöglicht die Berechnung sogenannter  $p$ -Werte. Diese werden als informales Maß für die Beweisstärke gegen die Nullhypothese beschrieben. Der Wert 0 steht für einen starken Beweis gegen die Nullhypothese, während wir bei dem Wert 1 wenig wissen. Je höher also ein  $p$ -Wert ist, desto weniger wissen wir – umso niedriger der  $p$ -Wert ist, desto stärker ist der Beweis, dass die Nullhypothese falsch ist. Mittels des Signifikanzniveaus  $\alpha$  und des sogenannten  $\alpha$  cut-offs wird anschließend die statistische Signifikanz eines Ergebnisses festgelegt. Der  $\alpha$  cut-off beschreibt eine Art Grenze, die bestimmt, ab welchem  $p$ -Wert ein Ergebnis als signifikant, bzw. nicht signifikant gilt (z.B.  $\alpha = 0.05$  – ist der  $p$ -Wert nun kleiner als 0.05, gilt das Ergebnis als signifikant).

Bei Nullhypothesensignifikanztests und  $p$ -Werten existieren jedoch diverse Probleme, die Dragicevic [Dra16] darlegt. Nullhypothesensignifikanztests stehen seit mehr als 50 Jahren in starker Kritik. Daher werden diese immer häufiger von Benutzer\*innen vermieden. Ein paar dieser Probleme werden im Folgenden genannt und sind der Grund, wieso auf eine Auswertung mit dieser Methode verzichtet wurde.

Nach Dragicevic [Dra16] ist es für Wissenschaftler\*innen, Lehrkräfte und Bücher üblich zu denken,  $p$  sei die Wahrscheinlichkeit, dass die Nullhypothese wahr ist. Dies wird von ihm als „fallacy of the transposed conditional“ beschrieben. Wie oben beschrieben steht ein  $p$ -Wert für die Beweisstärke gegen die Nullhypothese. Trotz der Darlegung, dass ein hoher  $p$ -Wert somit wenige Erkenntnisse zulässt, schlussfolgern einige Wissenschaftler\*innen, dass es daraufhin keinen Effekt gibt. Dragicevic beschreibt dies als gängigen Irrtum namens „accepting the null“, wodurch schon irreführende und falsche wissenschaftliche Schlussfolgerungen entstanden sind.

Nach der Arbeit von Dragicevic [Dra16] scheint es, dass nur wenige gute Argumente für die Darlegung von  $p$ -Werten existieren. Ein mögliches verbleibendes Argument ist, dass sich damit die statistische Signifikanz feststellen lässt. Jedoch stellt sich der binäre Signifikanztest (Signifikanz oder keine Signifikanz) als ungeeignet für faire statistische Kommunikation heraus. Nach dem  $\alpha = 0.05$  Kriterium erweist sich ein Ergebnis als

signifikant oder nicht signifikant. Somit kann angenommen werden, dass die signifikanten Ergebnisse den statistischen Test bestanden haben und für plausibel erklärt werden können. Andererseits werden nicht signifikante Ergebnisse so interpretiert, dass sie keinen Effekt oder keine Informationen haben, welches beides nicht korrekt ist. Dies und die dichotomische Denkweise (kategorisches Denken in Form von *Effekt* oder *kein Effekt*, *Beweis* oder *kein Beweis*, usw.) in der Statistik resultieren oftmals in möglichen Fehleinschätzungen.

Neben den genannten Kritikpunkten sind in der Arbeit von Dragicevic [Dra16] noch weitere Argumente gegen Nullhypothesensignifikanztests aufgeführt. In dieser Arbeit erfolgte die Auswertung daher mithilfe von Einschätzungen (*Estimation*) mittels Konfidenzintervallen. Anhand von Einschätzungen ist es beispielsweise möglich, Rankings zu definieren und die dichotomische Denkweise zu vermeiden.

## 2.6 Zusammenfassung

In diesem Kapitel wurden die für die vorliegende Arbeit wichtigsten Bereiche und Begriffe dargelegt. Diese umfassten den Begriff Visualisierung und verschiedene Visualisierungstechniken zur Darstellung von Zeitreihen. Zudem wurde erläutert, wie Informationen aus den Darstellungen gewonnen und verglichen werden können. Um die Ergebnisse dieser Studie verstehen zu können, wurden die statistischen Methoden zur Auswertung der Daten beschrieben. Zusätzlich wurden Argumente dargelegt, wieso wir auf die Auswertung durch Nullhypothesentests, die teilweise in anderen Arbeiten verwendet wurden, verzichteten.

## 3 Verwandte Arbeiten

Neben der vorliegenden Studie existieren weitere Arbeiten, die mithilfe von Studien ebenfalls diverse Visualisierungstechniken untersucht haben. Darunter fallen beispielsweise die Arbeiten von Heer et al [HKA09], Javed et al. [JME10] und Thudt et al. [TWP+16]. Dabei wurden bisher jedoch wenige Studien durchgeführt, die die Gemeinsamkeiten und Unterschiede von Split Space- und Shared Space-Techniken mit Hilfe von Linecharts, ThemeRiver und Small Multiples untersuchen.

Heer et al. [HKA09] verglichen in ihrer Arbeit Liniendiagramme und verschiedene Varianten des sogenannten *Horizon Graphs*. Dazu wurden zwei verschiedene Experimente durchgeführt. Das erste Experiment untersuchte den Einfluss der Horizon Graph-Varianten (*mirrored* und *offset*) und der Anzahl von Bändern innerhalb eines Horizon Graphs (zwei, drei und vier). Um dies zu untersuchen, wurden zwei Aufgaben festgelegt, die aufeinander aufbauten. Dabei wurden zwei Graphen gleichzeitig angezeigt, in denen jeweils genau ein Punkt markiert wurde, den es zu betrachten galt. Ziel der Aufgabe war es zu bestimmen, welcher Graph zum jeweiligen Punkt den größeren Wert hatte (*Discrimination*). Die zweite, aufbauende Aufgabe bestand darin, anschließend die Differenz der beiden Werte zu schätzen (*Estimation*). Dieses Experiment ergab, dass der Durchschnitt der Genauigkeit der ersten Aufgabe *Discrimination* für alle Bedingungen bei mindestens 99% lag. Daher wurde der Fokus auf die zweite Aufgabe gelegt. Die Ergebnisse dieser Aufgabe ergaben, dass die Fehlerrate bei vier Bändern im Horizon Graph höher ausfiel als die der anderen beiden Bandmengen. Zudem nahm die benötigte Zeit zum Lösen dieser Aufgabe mit der erhöhten Bandanzahl ebenfalls zu. Ein Unterschied zwischen den Horizon Graph Varianten konnte bei diesem Experiment nicht festgestellt werden.

Beim zweiten Experiment wurden Liniendiagramme und Horizon Graphs mit maximal zwei Bändern betrachtet. Dabei verwarfen Heer et al. die Variante *offset* und behielten ausschließlich die *mirrored* Variante des Horizon Graphs bei. Das Ziel dieses Experiments war es, die verschiedenen Visualisierungstechniken anhand verschiedener Diagrammhöhen zu vergleichen (Skalierungsfaktor: 1,  $\frac{1}{2}$ ,  $\frac{1}{4}$ ,  $\frac{1}{8}$  – Skalierungsfaktor 1 betrug dabei 48px). Die Aufgaben blieben die gleichen wie bei dem ersten Experiment. Dieses Experiment ergab, dass auch hier die durchschnittliche Genauigkeit für die erste Aufgabe bei mindestens 98% lag. Durch die sehr hohe Korrektheit bei der ersten Aufgabe wurde sich wieder auf die zweite Aufgabe konzentriert, welche ergab, dass die Fehlerrate bei abnehmender Diagrammhöhe zunahm. Zudem schnitt der Horizon Graph mit einem Band sowohl in der Genauigkeit, als auch in der Zeit besser ab als das Liniendiagramm. Somit lässt sich sagen, dass sich die Lesbarkeit mit Hilfe von Horizon Graphs bei kleineren Diagrammhöhen erhöhen lässt. Eine überraschende Erkenntnis war, dass die Teilnehmer\*innen die zweite Aufgabe schneller lösten, sobald die Diagrammgröße abnahm. Das bedeutet, dass kleinere Diagramme manchmal zu schnelleren Schätzungszeiten bei äquivalenten Fehlerraten führen. Der Unterschied zur vorliegenden Arbeit besteht darin, dass in dieser Arbeit die Unterschiede

zwischen Split Space- und zwei verschiedenen Shared Space-Techniken anhand von drei Unterscheidungsaufgaben verglichen werden. Zudem betrachtet diese Arbeit die Gegebenheit mehrerer Zeitreihen und befasst sich nicht mit dem Horizon Graph.

Die Arbeit von Javed et al. [JME10] ist eine der Arbeiten, die sowohl Split Space- als auch Shared Space-Techniken untersuchte. Der Hauptfokus bei dieser Arbeit lag darin zu untersuchen, ob komplexere Darstellungen in gewissen Situationen einen Vorteil gegenüber einfachen Liniendiagrammen haben. Dabei wurden Situationen mit limitiertem vertikalem Platz oder einer hohen Anzahl an Zeitreihen betrachtet. Dazu wurde eine Studie mit drei Aufgaben durchgeführt, in der vier verschiedene Visualisierungstechniken (Simple Line Graph, Braided Graph, Small Multiples und Horizon Graph) behandelt wurden. Die Aufgaben bestanden darin, die höchste Zeitreihe zu einem gewissen Punkt zu finden (*Maximum*), die Zeitreihe zu finden, die in einer Zeitspanne den höchsten Anstieg hat (*Slope*), und zu entscheiden, welche Zeitreihe zu zwei gegebenen Zeitpunkten größer ist (Zeitreihe  $A$  an Punkt  $P_A$  oder Zeitreihe  $B$  an Punkt  $P_B$ ) (*Discrimination*). Das weitere Studiendesign bezog drei verschiedene Zeitreihenmengen (zwei, vier und acht) und vertikale Diagrammgrößen (48px, 96px und 192px) ein. Die Studie ergab, dass sich der verfügbare vertikale Platz bei keiner der Visualisierungen auf die Zeit auswirkte. Für jede Größe ergab sich eine ähnliche Zeit. Umso stärker wirkte sich dieser jedoch negativ auf die Korrektheit der Ergebnisse aus. Javed et al. beschreiben dies als „classic time/accuracy trade-off“. Eine höhere Anzahl an Zeitreihen ergab eine niedrigere Korrektheit und eine erhöhte Antwortzeit. Mit Zunahme der Zeitreihenanzahl erhöhten sich die Unordnung und Überlappungen in den Shared Space-Techniken, während hingegen Split Space-Techniken gegenüber einer hohen Anzahl an Zeitreihen robuster waren. Insgesamt ergab sich, dass die Antwortzeiten des Liniendiagramms und der Small Multiples öfter schneller waren als die des Horizon und der Braided Graphs. Dies zeigt, dass das Liniendiagramm und die Small Multiples gegenüber verschiedenen Aufgaben robuster sind als die anderen beiden Visualisierungstechniken. Der Unterschied zu der Arbeit von Javed et al. liegt darin, dass in der vorliegenden Arbeit die Unterschiede der Visualisierungstechniken nicht anhand variierender Diagrammgrößen ausgemacht werden. Obwohl eine der Aufgaben der Arbeit von Javed et al. auch in der vorliegenden Arbeit zu finden ist, unterscheiden sich die restlichen beiden Aufgaben insofern, dass in der vorliegenden Arbeit aggregierte Werte und Werte über einen Zeitraum hinweg betrachtet und verglichen werden müssen. Zudem wurde hier eine Visualisierung untersucht, deren Basislinie nicht der X-Achse eines Diagramms entspricht.

Des Weiteren untersuchten Thudt et al. [TWP+16] die Lesbarkeit von geschichteten Graphen. Dabei wurde insbesondere der Einfluss von statischen und interaktiven Basislinien, Symmetrie und Wiggle-Faktor auf die Extraktion und das Vergleichen von sowohl individuellen als auch aggregierten Werten und der Lesbarkeit von Trends untersucht. Dazu wurde eine Studie durchgeführt, deren Design vier verschiedene geschichtete Visualisierungsarten und drei verschiedene Aufgaben umfasste. Die Visualisierungsarten bestanden aus dem Stacked Area Chart, dem ThemeRiver, dem Streamgraph und dem von Thudt et al. vorgestelltem ThemeRiver mit interaktiver Basislinien-Begradigung. Bei Letzterem konnte eine Zeitreihe angewählt werden, wodurch anschließend die Basislinie so gestaltet wurde, dass die ausgewählte Zeitreihe mittig oberhalb der Basislinie angezeigt wurde. Die erste Aufgabe bestand aus einem direkten Vergleich, bei dem zwei verschiedene Werte zweier unterschiedlicher Zeitreihen miteinander verglichen werden mussten. Anschließend musste entschieden werden, welcher der Werte größer ist. Die zweite Aufgabe umfass-

te den Vergleich von Streams. Hierbei wurde eine Zeitreihe in Form eines Area Charts angezeigt. Anschließend musste entschieden werden, welche Zeitreihe der geschichteten Visualisierungen der angezeigten Zeitreihe im Area Chart entspricht. Die dritte Aufgabe bestand aus einem globalen Vergleich. Hierbei wurden zwei Punkte in einem Diagramm markiert. Anschließend mussten die Teilnehmer\*innen entscheiden, an welchem Punkt die Kombination, bzw. Aggregation, der jeweiligen Zeitreihenwerte größer ist. Die Studie ergab, dass die Performanz der Visualisierungen stark von der zu lösenden Aufgabe abhängt. Der Streamgraph führte jedoch bei den Vergleichen von individuellen und aggregierten Werten zu einer besseren Performanz als die anderen beiden statischen Visualisierungen ThemeRiver und Stacked Area Chart. Die von Thudt et al. vorgestellte interaktive Variante des ThemeRivers lieferte jedoch bessere Ergebnisse bei Vergleichen von Streams. Die vorliegende Arbeit unterscheidet sich von der Arbeit von Thudt et al. insofern, dass sie nicht nur Shared Space-Techniken betrachtet, sondern auch Split Space-Techniken und diese sowohl untereinander als auch miteinander vergleicht. Der Einbezug von interaktiven Visualisierungstechniken findet hierbei nicht statt. Thudt et al. verwendeten in ihrer Arbeit zudem deutlich höhere Zeitreihenmengen (10, 30, 300) als wir.

Ein wesentlicher Unterschied der vorliegenden zu den drei genannten Arbeiten besteht darin, dass diese allesamt vor Ort durchgeführt wurden. Die Benutzerstudie dieser Arbeit hingegen wurde online durchgeführt.



## 4 Benutzerstudie

Die Benutzerstudie basiert auf einem Studiendesign, mit welchem die Lesbarkeit der drei verschiedenen Visualisierungstechniken mit mehreren Zeitreihen untersucht werden soll. Das Studiendesign ist darauf ausgelegt, zu messen, wie sich Aufgabenstellung, Visualisierungstechnik und Zeitreihenanzahl auf die jeweils benötigte Antwortzeit und Korrektheit auswirken. Der genauere Aufbau ist in den folgenden Unterkapiteln dargelegt.

### 4.1 Aufgaben

Wie schon erwähnt, nehmen wir an, dass geschichtete Graphen, Liniendiagramme und Small Multiples verschiedene Stärken und Schwächen aufweisen. Daher wurden Aufgaben definiert, die beim Gebrauch von Visualisierungen mit Zeitreihen Anwendung finden können und deren Ergebnisse auch messbar sind. Zwar ist es möglich, dass diese Aufgaben in der realen Welt auf diese Weise nicht vorkommen, jedoch nutzten auch andere Arbeiten synthetische bzw. künstliche Aufgaben zur Durchführung von Studien [HKA09; JME10; TWP+16]. Bei der Durchführung von synthetischen Aufgaben müssen die Teilnehmer\*innen ein Diagramm mit einer zielgerichteten Betrachtungsweise untersuchen, sodass sie die jeweilige Aufgabe lösen können. Stellte sich die Antwort der Teilnehmer\*innen als korrekt heraus, so kann man davon ausgehen, dass das Diagramm und die Aufgabe verstanden wurde. Durch die korrekte Bearbeitung von synthetischen Aufgaben hoffen wir, Aussagen über reale Aufgaben machen zu können.

#### 4.1.1 Aufgabe 1: Vergleich von mehreren Werten (Maximum)

Bei dieser Aufgabe sollten die Teilnehmer\*innen entscheiden, welche Zeitreihe zum angezeigten Zeitpunkt A den größten Wert hat. Falls mehrere Zeitreihen zufälligerweise den gleichen höchsten Wert hatten, gab es mehrere richtige Antwortmöglichkeiten. Diese Aufgabe forderte die Teilnehmer\*innen dazu auf, die verschiedenen Zeitreihen an einem lokalen Punkt direkt miteinander zu vergleichen. In Abbildung 4.1 ist diese Aufgabe beispielhaft dargestellt.

Hierbei kann man von einer Basisaufgabe oder leichten Aufgabe sprechen, da die Teilnehmer\*innen bei dieser Aufgabe die Werte der einzelnen Zeitreihen betrachten mussten. Die Teilnehmer\*innen mussten anschließend in der Lage sein, die einzelnen Werte vergleichen zu können. Dies beschreibt eine Standardaufgabe für die Evaluation von Zeitreihenvisualisierungen [TWP+16].

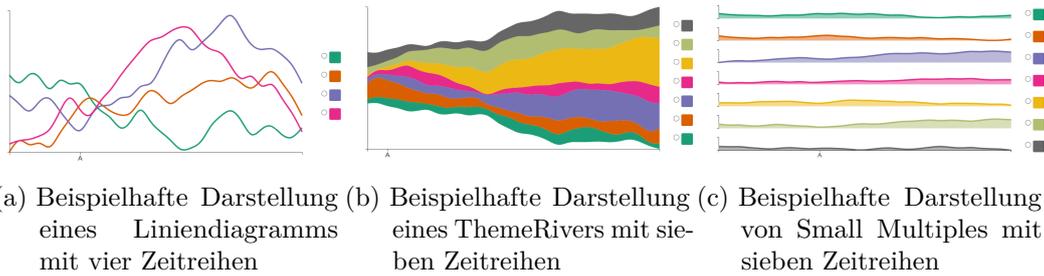


Abbildung 4.1: Beispielhafte Darstellungen der Aufgabe Maximum anhand der verschiedenen Visualisierungstechniken.

#### 4.1.2 Aufgabe 2: Vergleich von mehreren Flächen (Area)

Die Teilnehmer\*innen sollten bei dieser Aufgabe entscheiden, welches Integral bzw. welche Fläche unter den angezeigten Zeitreihen im markierten Zeitraum (von A bis B) am größten ist. Falls mehrere Flächen zufälligerweise gleich groß waren, gab es mehrere richtige Antwortmöglichkeiten. Für die Aufgabe sind in Abbildung 4.2 Beispiele zu sehen.

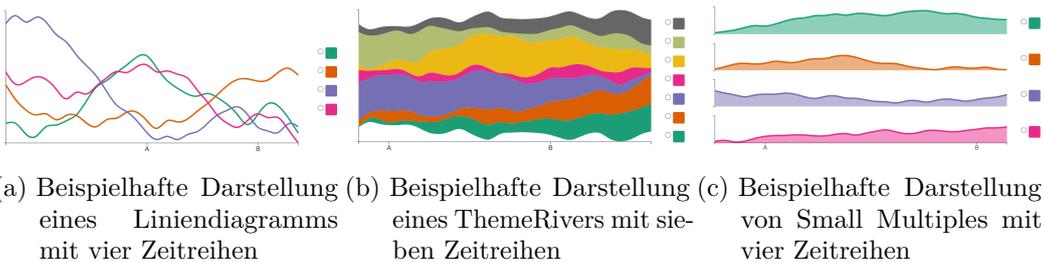


Abbildung 4.2: Beispielhafte Darstellungen der Aufgabe Area anhand der verschiedenen Visualisierungstechniken.

Hierbei mussten die Teilnehmer\*innen nicht nur Werte an einem Punkt lesen und vergleichen können, sondern die einzelnen Werte einer Zeitreihe visuell aggregieren und die Summen der Zeitreihenwerte anschließend vergleichen können. Durch das Aggregieren der einzelnen Werte und Vergleichen der Flächen wird diese Aufgabe als schwieriger eingeordnet als das Vergleichen von einzelnen Werten an einem Punkt.

#### 4.1.3 Aufgabe 3: Vergleich von Summen (Global)

Hierbei sollten die Teilnehmer\*innen die jeweiligen Zeitreihenwerte zu den gegebenen Zeitpunkten A oder B aufsummieren und entscheiden, an welchem der Punkte diese Summe größer ist. Falls die Summe der Werte gleich war, galten beide Antworten als richtig. Eine beispielhafte Veranschaulichung ist in Abbildung 4.3 zu sehen.

Hierbei handelt es sich um eine Aufgabe, die einen globalen Vergleich benötigt. Diese Aufgabe wird ebenso als leicht eingestuft, da hierbei die Teilnehmer\*innen die einzelnen Werte an den markierten Punkten betrachten und visuell aggregieren sollten. Diese Aggregationen

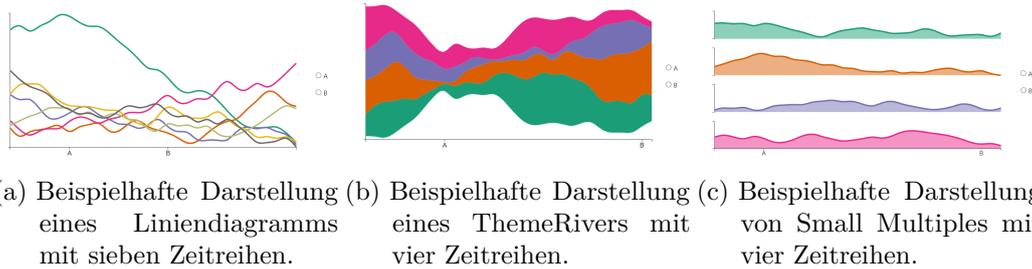


Abbildung 4.3: Beispielhafte Darstellungen der Aufgabe Global anhand der verschiedenen Visualisierungstechniken.

sollen anschließend direkt miteinander verglichen werden. Auch diese Aufgabenart wurde des Öfteren für die Evaluation von Zeitreihenvisualisierungen genutzt [TWP+16].

## 4.2 Erwartete Effekte

Wir nahmen an, dass die verschiedenen Visualisierungstechniken unterschiedliche Vor- und Nachteile haben, die sich je nach Art der Aufgabe auf die Antwortzeit und die Korrektheit der Antworten auswirken. Daher erwarteten wir folgende Effekte:

- E1 Die Antwortzeit für die Aufgabe Maximum ist geringer als die Antwortzeit der anderen beiden Aufgaben.
- E2 Für die Aufgabe Maximum ist die Antwortzeit für das Liniendiagramm kürzer als die Antwortzeit der anderen beiden Visualisierungstechniken.
- E3 Für die Aufgabe Maximum ist die Korrektheit für das Liniendiagramm höher als die Korrektheit für den ThemeRiver und diese ist höher als die Korrektheit für die Small Multiples.
- E4 Für die Aufgabe Area ist die Antwortzeit für den ThemeRiver kürzer als die Antwortzeit der anderen beiden Visualisierungstechniken.
- E5 Für die Aufgabe Area ist die Korrektheit für den ThemeRiver höher als die Korrektheit der anderen beiden Visualisierungstechniken.
- E6 Für die Aufgabe Global ist die Antwortzeit für den ThemeRiver kürzer als die Antwortzeit der anderen beiden Visualisierungstechniken.

E1 formulierten wir, da es sich bei der Aufgabe Maximum um eine lokale Aufgabe mit einem markierten Punkt handelt. Dabei müssen lediglich direkte Vergleiche zwischen den verschiedenen Zeitreihen zu einem Zeitpunkt getätigt werden. Die Betrachtung von Zeitreihen an nur einem Punkt kostet dabei nicht so viel Zeit wie die Betrachtung der Zeitreihen an oder zwischen zwei Punkten.

Wir haben E2 und E3 formuliert, da es in einem Liniendiagramm naheliegt, die einzelnen Zeitreihen leicht miteinander vergleichen zu können. Der Vergleich zwischen Zeitreihen in einer Shared Space-Technik wurde dabei als einfacher und somit schneller eingestuft, als der

Vergleich zwischen Zeitreihen in einer Split Space-Technik. Zudem ist das Liniendiagramm eines der gängigsten Diagramme im Alltag. Durch gewisse Gewohnheiten sollten die Teilnehmer\*innen mit dieser Visualisierungstechnik besser zurechtkommen. Daher erwarteten wir, dass das Diagramm für diese Art von Aufgaben eine kürzere Antwortzeit hat und zudem eine höhere Korrektheit aufweisen wird.

E4 und E5 erwarteten wir, da in einem ThemeRiver die Zeitreihen in Form von Flächen übereinandergeschichtet dargestellt werden. Somit existiert in dieser Visualisierungstechnik der Faktor Unordnung nicht. Zudem sollte der Vergleich der Flächen durch die Darstellung in einem gemeinsamen Diagramm besser funktionieren als in einer Shared Space-Technik. Anhand dessen sollten die Teilnehmer\*innen die Flächen leicht analysieren und vergleichen können.

Der Effekt E6 wurde erwartet, da in einem ThemeRiver alle Zeitreihen bereits übereinander liegen. Anhand dessen muss im Gegensatz zu den anderen Visualisierungstechniken keine weitere gedachte visuelle Schichtung der Zeitreihen erfolgen, deren Wahrnehmung zusätzlich Zeit kosten würde. Daher können in einem ThemeRiver die Summen der Zeitreihen zu den markierten Zeitpunkten schneller verglichen werden.

Um den Einfluss der Anzahl der Zeitreihen auf die Lesbarkeit zu untersuchen, wurden unabhängig von der Visualisierungstechnik folgende Effekte erwartet:

E7 Für die Aufgabe Area ist die Antwortzeit für vier Zeitreihen kürzer als die Antwortzeit für sieben Zeitreihen.

E8 Für die Aufgabe Area ist die Korrektheit für vier Zeitreihen höher als die Korrektheit für sieben Zeitreihen.

Wir formulierten E7, da der Vergleich von mehreren Zeitreihen mehr Zeit in Anspruch nimmt und diese dabei insbesondere über einen Zeitraum hinweg betrachtet werden müssen. Zudem ist es in einem Liniendiagramm zusätzlich möglich, dass bei einer höheren Anzahl an Zeitreihen die Unordnung zunimmt. Somit kann es länger dauern, den jeweiligen Zeitreihen aufmerksam zu folgen und diese anschließend zu vergleichen. Besonders durch die genannte Unordnung ist es möglich, dass die Teilnehmer\*innen die Zeitreihen nicht mehr so gut erkennen und differenzieren können. Insbesondere bei der Aufgabe Area ist dies von Wichtigkeit, da dabei der Bereich unterhalb der Zeitreihe ausschlaggebend ist. Da dies in diesem Fall, neben der erhöhten Vergleichsanzahl, zu einer erhöhten Fehlerwahrscheinlichkeit führen kann, wurde E8 formuliert.

### 4.3 Teilnehmer\*innen

Insgesamt nahmen 48 Teilnehmer\*innen an der Studie teil, die allesamt von der Crowdsourcing Plattform *Prolific* stammten. Dabei entschieden wir uns durch die Möglichkeit der Vorsortierung über *Prolific*, ausschließlich Teilnehmer\*innen zuzulassen, die die englische Sprache flüssig sprechen konnten. Alle Teilnehmer\*innen nahmen freiwillig teil und wurden nach erfolgreicher Teilnahme der Studie mit ungefähr 6 € (11,16 € pro Stunde) bezahlt. Die Teilnehmer\*innen waren zwischen 18 und 50 Jahren alt und in verschiedenen Berufsfeldern tätig. Zusätzlich stellten wir bei 14 von 48 Teilnehmer\*innen eine Rot-Grün-Sehschwäche

fest. Die genannten und auch weitere Verteilungen der demographischen Daten sind im Anhang unter Kapitel A.3 in den Abbildungen A.15 bis A.18 zu finden.

## 4.4 Apparat

Aufgrund der online durchgeführten Studie konnte die Einheitlichkeit der genutzten Geräte nicht sichergestellt werden, da die Teilnehmer\*innen ihre eigenen Geräte nutzten. Dabei wurden die Teilnehmer\*innen gebeten, die Studie möglichst an einem Gerät mit einer Maus durchzuführen und von der Nutzung von mobilen Geräten wie Smartphones oder Tablets abzusehen. Da davon auszugehen war, dass die Teilnehmer\*innen Monitore mit verschiedenen Auflösungen und Größen nutzten, mussten die Teilnehmer\*innen vor dem eigentlichen Beginn der Studie eine kurze Skalierungsaufgabe absolvieren. Diese erfolgte, indem die Teilnehmer\*innen gebeten wurden, eine Kreditkarte (oder eine Karte ähnlicher Größe) zur Hand zu nehmen und ein angezeigtes Rechteck einmalig auf die Größe der Karte zu ziehen. Anhand der Größe des Rechtecks wurden die Diagramme daraufhin skaliert. Auf die technisch bedingte Farbdarstellung und die Helligkeit der einzelnen Monitore konnte kein Einfluss genommen werden. Der Abstand zum Monitor konnte unter den Teilnehmer\*innen ebenso variieren. Das Verschaffen anderweitiger Vorteile, z.B. das Heranzoomen im Browser durch ein Tastenkürzel, konnte nicht ausgeschlossen werden. Jedoch wurden solche Gegebenheiten aufgezeichnet und herausgefiltert. Dies kam in der gesamten Studie ein Mal vor.

## 4.5 Szenario

Wir führten die Benutzerstudie ausschließlich auf Englisch durch, da die englische Sprache deutlich weiterverbreitet ist als die deutsche Sprache. Zudem liegt der prozentuale Anteil der englischsprachigen Benutzer\*innen von Prolific deutlich über dem Anteil der deutschsprachigen Benutzer\*innen [Pro14]. Anhand dessen erwarteten wir eine bessere Stichprobe für die Studie.

Pro Versuch wurde den Teilnehmer\*innen ein Diagramm mit einer bestimmten Visualisierungstechnik und einer kontrollierten Anzahl von Zeitreihen gezeigt. Die Größe des Diagramms hing von der Skalierung des Rechtecks ab und variierte unter den Teilnehmer\*innen. Die Zeitreihen wurden nach dem Farbschema *Dark2* des *ColorBrewers2* [HB] eingefärbt. Dabei trug jede Zeitreihe eine festgelegte Farbe.

Die Antworten wurden über Radiobuttons gegeben, die in derselben Farbe eingefärbt waren wie die dazugehörige Zeitreihe. Diese konnte man nur mit der Maus anklicken. Sobald einer dieser Radiobuttons angewählt wurde, wurde die Antwort übermittelt. Die Zeit zum Bewegen des Mauscurors wurde dabei in Kauf genommen. Von der Auswahl der Radiobuttons mittels der Nummern einer Tastatur wurde abgesehen, da davon ausgegangen wurde, dass manche Teilnehmer\*innen nach dem Betrachten des Diagramms zusätzlich noch auf die Tastatur schauen müssten, um die richtige Zahl auszuwählen. Da dies zusätzliche Zeit kosten würde, die die Zeit des Bewegens des Mauscurors übertreffen würde, konnte man die Radiobuttons nur mit der Maus auswählen. Zudem schätzten wir die Wahrscheinlichkeit

des versehentlichen Verklickens auf der Tastatur höher ein als das versehentliche Anklicken eines falschen Radiobuttons mittels der Maus.

Interaktionsmöglichkeiten mit einem Diagramm waren nicht gegeben. Die einzigen beiden Interaktionsmöglichkeiten während der Studie bestanden aus dem Anwählen der Radiobuttons und dem Skalieren des Rechtecks zu Beginn.

Die Auswahl der Daten, die mittels der Visualisierungstechniken angezeigt wurden, geschah zufällig. Dabei wurde sichergestellt, dass pro Aufgabe keine doppelten Teildatensätze vorkamen. Der gesamte Datensatz setzte sich aus generierten Teildatensätzen zusammen. Weitere Details sind in Kapitel 4.6 zu finden.

Die Studie war in Blöcke unterteilt. Ein Block bestand dabei aus einer Aufgabe (Maximum, Area, Global), deren szenischer Beschreibung und den dargestellten Diagrammen. Insgesamt existierten also drei Blöcke, deren Reihenfolge randomisiert war. Um mögliche Lerneffekte im besten Fall zu vermeiden, wurde die Reihenfolge der angezeigten Visualisierungsart sowie die Anzahl der Zeitreihen für alle Teilnehmer\*innen ebenso zufällig gewählt. Details zu den Visualisierungsarten und den Zeitreihen werden in den Kapiteln 4.9.2 und 4.9.3 beschrieben.

## 4.6 Datensätze

Für die Erstellung des Datensatzes fertigten wir mittels einer *Random Walk*-Implementation verschiedene zufällig generierte Teildatensätze an. Dabei lief der Random Walk eine festgelegte Anzahl von Schritten. Die Anzahl der Schritte betrug für alle Datensätze 30 Zeitpunkte. Bei der Wahl der Anzahl der Zeitpunkte orientierten wir uns an Thudt et al. [TWP+16], die für ihre randomisierten Daten ebenso 30 Zeitpunkte wählten. Um die generierten Graphen realitätsnäher zu gestalten, wurde die jeweilige Schrittgröße mit Hilfe eines *Stepsets* bestimmt. Innerhalb des Stepsets befanden sich in dieser Studie die Zahlen 0 bis 5. Von diesen Zahlen wurde pro Schritt eine zufällig bestimmt, welche dann die Schrittgröße beschrieb. Der Random Walk konnte dadurch auch im Wertebereich von 0 bis 5 auf der Y-Achse beginnen und war nicht zwingenderweise an den Punkt (0,0) gebunden. Ein Schritt eines Random Walks erfolgte dabei mit einer Wahrscheinlichkeit von 50% nach oben oder nach unten. Ein Ausschnitt der Random Walk-Implementation ist in Form eines Pseudocodes in Algorithmus 4.1 zu sehen.

Um negative Werte innerhalb einer Zeitreihe zu verhindern, die Einheitlichkeit des Aufbaus der Zeitreihen jedoch beizubehalten, wurde anschließend ein Normierungsverfahren angewandt. Dieses ist in Algorithmus 4.2 durch einen Pseudocode dargestellt. Das Vorgehen bestand darin, dass der niedrigste Wert einer Zeitreihe im Falle einer negativen Ausprägung auf alle Werte dieser Zeitreihe aufsummiert wurde, inklusive eines zufälligen Offsets zwischen 0 und 5. Im Falle, dass der niedrigste Wert einer Zeitreihe im positiven Bereich lag, wurde dieser von allen Werten der Zeitreihe abgezogen und wiederum ein zufälliges Offset zwischen 0 und 5 aufaddiert. Durch die Normierung wurden negative Werte verhindert und die Vergleichbarkeit blieb erhalten. Die Normierung der Schritte ist in Algorithmus 4.2 durch einen Pseudocode beschrieben.

---

**Algorithmus 4.1** Generierung der Random Walk-Schritte.

---

```
1: procedure SCHRITTGENERATOR( $N$ )
2:    $stepset \leftarrow [0, 1, \dots, 5]$ 
3:   for  $i \leftarrow 0$  to  $N$  do
4:      $step \leftarrow$  wähle zufälligen Wert aus  $stepset$ 
5:      $direction \leftarrow$  wähle zufällig entweder 0 oder 1
6:     if  $direction$  is 1 then
7:        $walk(i + 1) \leftarrow walk(i) + step$ 
8:     else
9:        $walk(i + 1) \leftarrow walk(i) - step$ 
10:    end if
11:  end for
12: end procedure
```

---

---

**Algorithmus 4.2** Normierung der Random Walk-Schritte.

---

```
1: procedure NORMALISIERUNG( $walk$ )
2:    $N \leftarrow$  Schritte in  $walk$ 
3:    $minimum \leftarrow \min(walk)$ 
4:    $offset \leftarrow$  wähle zufälligen Wert zwischen 0 und 5
5:   for  $i \leftarrow 0$  to  $N$  do
6:      $walk(i) \leftarrow walk(i) - minimum$ 
7:      $walk(i) \leftarrow walk(i) + offset$ 
8:   end for
9: end procedure
```

---

Insgesamt bestanden die verwendeten Daten aus 120 verschiedenen Teildatensätzen. Zudem waren von den 120 Teildatensätzen 60 für vier Zeitreihen und 60 für sieben Zeitreihen vorgesehen.

Thudt et al. [TWP+16] nutzten in ihrer Studie einen Teildatensatz von Daten, die die New York Times in ihrer Streamgraph-Visualisierung namens "The Ebb and Flow of Box Office Sales" [BBCC08] verwendete. Dabei handelt es sich um einen Datensatz mit dem Brutto-Einkommen verschiedener Filme [Gra]. Thudt et al. verwendeten für ihre Darstellung der Daten insgesamt 300 Zeitreihen gleichzeitig. Der Versuch der Nutzung von nur vier oder sieben Zeitreihen mit einem Teildatensatz der realen Daten ergab für jede der Visualisierungsart eine periodische Darstellung der Zeitreihen. Da die Wahl der markierten Punkte A und B für die Aufgaben zufällig erfolgte, war es möglich, dass einer der Punkte in der Nähe eines sehr niedrigen Wertes lag und der andere Punkt in der Nähe eines sehr hohen Wertes. Dadurch konnte eine hohe Differenz zwischen den Werten existieren, wodurch eine sehr einfache Unterscheidung der Summen der Zeitreihen möglich war. Deshalb konnte die Aufgabe Global nicht sinngemäß gelöst werden. Zusätzlich unterschieden sich die Zeitreihen meist stark, sodass die Aufgaben generell nicht sinngemäß durchgeführt werden konnten. Wir versuchten zudem, vergleichbare Zeitreihen in einen Teildatensatz zu integrieren, um große Unterschiede zwischen den Zeitreihen zu vermeiden. Da wir jedoch dabei bewusst die Daten wählten und dies nicht im Sinne der Studie lag, entschieden wir uns gegen die Nutzung eines Teildatensatzes der Filmeinkommen.

## 4.7 Studienablauf

Die Teilnehmer\*innen wurden von uns gebeten, die vorneweg angezeigten Anweisungen und Informationen aufmerksam zu lesen. In diesen wurde den Teilnehmer\*innen eine Einleitung dargelegt und darauf hingewiesen, sich in eine normale Sitzposition zu bringen. Um eine möglichst gleiche Ausgangssituation zu schaffen und unfaire Vorteile zu verhindern, sollten die Teilnehmer\*innen keine externen Hilfsmittel wie z.B. einen Stift oder die Hände hinzuziehen und sich auch nicht mit dem Kopf dem Bildschirm nähern.

Nach den Instruktionen folgte für die Teilnehmer\*innen ein demographischer Fragebogen (siehe Abbildung A.2) und eine kurze Kalibrierungsaufgabe (siehe Abbildung A.3), die für die Skalierung der Diagramme nötig war (siehe Kapitel 4.4). Um zu testen, ob sich unter den Teilnehmer\*innen Personen befanden, die eine Farbsehschwäche besitzen, mussten die Teilnehmer\*innen einen kurzen Farbsehtest absolvieren. Hierbei wurde allen Teilnehmer\*innen dieselbe Farbtafel (siehe Kapitel 4.9.1) angezeigt.

Anschließend gelangten die Teilnehmer\*innen zur ersten der drei Aufgaben. Für jede Aufgabe wurde ein exemplarisches Szenario präsentiert, das den Teilnehmer\*innen helfen sollte, den Nutzen des Diagramms und die Aufgabe näher zu bringen. Um Übungseffekte zu reduzieren, mussten die Teilnehmer\*innen vor jeder Aufgabe mindestens einen Übungsdurchlauf absolvieren, in dem sie die jeweilige Aufgabe anhand der Diagramme absolvierten. Jede Visualisierungstechnik wurde hier genau ein Mal mit je einer der beiden Mengen der Zeitreihen angezeigt. Insgesamt bestand ein Übungsdurchlauf somit aus sechs Diagrammen. Sobald die Teilnehmer\*innen auf den Start-Knopf drückten, wurde das erste Diagramm angezeigt. Sobald eine Antwort mittels eines Radiobuttons ausgewählt wurde, wurde das Diagramm ausgeblendet. Nach einer Pause von zwei Sekunden erschien das nächste Diagramm. Während des Übungsdurchlaufs bekamen die Teilnehmer\*innen eine Rückmeldung, ob ihre gegebene Antwort korrekt oder falsch gewesen ist. Nach dem Übungsdurchlauf konnten die Teilnehmer\*innen wählen, ob sie den Hauptdurchgang beginnen oder nochmals einen Übungsdurchgang absolvieren wollen.

Der jeweilige Hauptdurchgang verlief analog zu einem Übungsdurchgang, jedoch mit einer höheren Anzahl an Diagrammen und ohne Rückmeldung über die Korrektheit einer gegebenen Antwort. Während des Hauptdurchlaufs wurde den Teilnehmer\*innen ein Fortschrittbalken angezeigt, auf dem der Gesamtfortschritt der Studie durch eine Prozentzahl angezeigt wurde. Sobald die Teilnehmer\*innen mit einer Aufgabe fertig waren, wurde die Beschreibung, inklusive eines Szenarios, für die nächste Aufgabe angezeigt. Zwischen den jeweiligen Aufgaben hatten die Teilnehmer\*innen die Möglichkeit, eine Pause einzulegen, falls diese benötigt wurde. Während der Bearbeitung einer Aufgabe waren keine längeren Pausen vorgesehen.

Nach Beendigung der drei Aufgaben wurden die Teilnehmer\*innen gebeten, anzugeben, welche der Visualisierungstechniken sie am meisten und welche sie am wenigsten mochten. Dabei bestand auch die Möglichkeit, keine bestimmte Visualisierungstechnik anzugeben. Nachdem die Teilnehmer\*innen ihre Antworten abgeschickt hatten, wurde diesen eine Danksagung und ein benötigter Abschlusscode angezeigt.

Ausschnitte eines beispielhaften Ablaufs der Studie sind in Kapitel A.2 im Anhang zu finden.

## 4.8 Studiendesign

Die Studie wurde mit den folgenden Faktoren als ein *Within-subjects*-Experiment entworfen:

- **Anzahl Zeitreihen (Z):** 4, 7
- **Visualisierungstyp (V):** Liniendiagramm (LC), Small Multiples (SM), ThemeRiver (TR)
- **Aufgaben (A):** Maximum, Area, Global

Durch die Faktoren entstand ein  $Z \times V \times A$  Design mit  $2 \times 3 \times 3 = 18$  verschiedenen Bedingungen für alle Teilnehmer\*innen. Jede Bedingung wurde fünf Mal wiederholt, um ein präziseres und robusteres Ergebnis zu erhalten. Daraus resultierten 90 Versuche pro Teilnehmer\*in während der Hauptdurchläufe. Inklusive der Übungsdurchläufe, in denen jede Bedingung ein Mal wiederholt wurde, beläuft sich die Gesamtzahl der Versuche auf mindestens 108. Die Gesamtzahl der Versuche konnte pro Teilnehmer\*in höher ausfallen, da die Möglichkeit bestand, einen Übungsdurchlauf zu wiederholen.

Die Reihenfolge der Aufgaben geschah mittels einfacher Randomisierung und war somit zufällig, ebenso wie die Reihenfolge der Visualisierungstypen innerhalb einer Aufgabe und die Anzahl der Zeitreihen.

Wir generierten für die Studie mehr Teildatensätze als nötig. Dadurch und durch die Randomisierungen bekamen die Teilnehmer\*innen mit einer hohen Wahrscheinlichkeit verschiedene Daten angezeigt. Eine Überlappung der Daten zwischen den Teilnehmer\*innen konnte nicht ausgeschlossen werden, wurde jedoch nicht weiter überprüft.

## 4.9 Experimentelle Bedingungen

Für die bestmögliche Durchführung der Benutzerstudie führten wir zwei Vorstudien durch, die in Kapitel 5 genauer beschrieben werden. Die Ergebnisse und das Feedback der Teilnehmer\*innen der beiden Vorstudien nutzten wir, um die finale Version der Benutzerstudie zu entwerfen. Die verschiedenen Designentscheidungen werden in den folgenden Punkten genauer erläutert.

**Aufgabentypen** Die Teilnehmer\*innen der Vorstudien gaben an, die verschiedenen Aufgaben direkt verstanden zu haben. Falls Unklarheiten auftraten, wurden diese mit Hilfe der Übungsdurchgänge beseitigt. Zudem wurden die szenischen Beschreibungen der Aufgaben als hilfreich und nicht störend erachtet.

### 4.9.1 Farbsehtest und Farbwahl

Der Farbsehtest bestand aus einer *Ishihara-Tafel* mit einer Zahl (siehe Abbildung 4.4) und prüfte ausschließlich auf eine Rot-Grün-Sehschwäche. Von einer Prüfung der Gelb-Blau-Sehschwäche oder einer vollständigen Farbenblindheit wurde abgesehen, da diese lediglich mit einer Häufigkeit von 1:100.000 auftreten [Koh13; Orp07]. Die Vorstudien ergaben, dass der Farbtest die Teilnehmer\*innen mit einer Rot-Grün-Schwäche mit sehr hoher Wahrscheinlichkeit ausfindig macht. Jedoch waren durch das Erraten der Zahl durch die Teilnehmer\*innen falsch-negative Ergebnisse nicht auszuschließen. Ebenso war nicht auszuschließen, dass es Teilnehmer\*innen gab, die aufgrund von Gleichgültigkeit irgendeine Zahl in das Antwortfeld eintrugen. Somit waren auch falsch-positive Ergebnisse möglich.

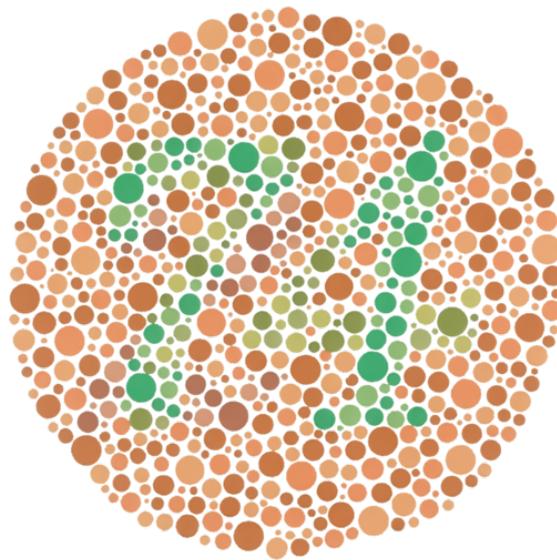


Abbildung 4.4: Darstellung der Ishihara-Tafel, die in der Studie zur Prüfung der Rot-Grün-Schwäche verwendet wurde. Personen mit einer Rot-Grün-Schwäche können hierbei eine 21 sehen. Normalsichtige Personen sehen eine 74 (die Zahl 71 wurde auch akzeptiert).

Die Wahl der Farbpalette *Dark2* für die Zeitreihen wurde mittels *ColorBrewer2* [HB] anhand der Empfehlung von Harrower und Brewer [HB03] getroffen. Javed et al. [JME10] nutzten in ihrer Studie ebenfalls diese Farbpalette mit der Begründung einer Gewährleistung der graphischen Wahrnehmung jeder Zeitreihe. Weitere Justierungen nahmen wir anhand des Feedbacks der Vorstudien vor. In der ersten Vorstudie stellte sich eine der Farben als problematisch heraus und wurde durch eine andere Farbe der Farbpalette *Dark2* ersetzt (siehe Kapitel 5.2.1).

Nachdem in der zweiten Vorstudie weitere Farberkennungsprobleme – insbesondere bei der Erkennung zweier grünlastiger Linien im Liniendiagramm – gemeldet wurden, wurden diese Farben diskutiert und durch zwei farbenblind-freundliche Farben des ColorBrewers ersetzt. Da es sich bei den beiden grünlastigen Farben angeblich ebenfalls um farbenblind-freundliche Farben handelte, konnte weiterhin eine verfälschte Wahrnehmung durch eine Rot-Grün-Sehschwäche nicht komplett ausgeschlossen werden. Da sich in unserem Team

eine Person mit einer Rot-Grün-Sehschwäche befand, testeten wir die gesamte Farbpalette anschließend informell und empfanden sie als geeignet. Die genutzte Farbpalette ist in Abbildung A.21 im Anhang zu finden.

## 4.9.2 Visualisierungstechniken

Um zu verhindern, dass die Teilnehmer\*innen mathematische Berechnungen oder Vergleiche anhand der Achsenbeschriftungen vornehmen, wurden die Beschriftungen an den Achsen jeder Visualisierungstechnik entfernt. Wichtig ist ebenso, dass jedes Diagramm die gleiche Höhe und Breite besaß. Die folgenden Visualisierungstechniken, welche in Kapitel 2.2 detaillierter erklärt wurden, wurden für die Vorstudien und die richtige Studie verwendet:

- **Liniendiagramm/Linechart (LC):** Eine grundlegende Visualisierungstechnik, in der mehrere Zeitreihen gemeinsam in einem Diagramm angezeigt wurden. Die Linien wurden zur besseren Unterscheidung mit einer Farbe versehen.
- **Small Multiples (SM):** Mehrere einzelne Diagramme, die vertikal untereinander angeordnet wurden. Dabei wurde jede Zeitreihe in einem einzelnen Diagramm angezeigt. Die resultierenden Linien und die jeweils darunterliegende Fläche wurden zur besseren Unterscheidung mit einer Farbe versehen.
- **ThemeRiver (TR):** Eine Visualisierungstechnik, bei der mehrere übereinander geschichtete Zeitreihen um eine zentrale Achse angezeigt werden. Daraus resultiert pro Zeitreihe eine visuelle Fläche, die zur besseren Unterscheidung mit einer Farbe versehen wurde.

## 4.9.3 Anzahl der Zeitreihen

Heer et al. [HKA09] beschränkten sich in ihrer Publikation auf lediglich zwei Zeitreihen, während Javed et al. [JME10] in ihrer Studie zwei, vier und acht Zeitreihen nutzten. Um den Einfluss der Anzahl von Zeitreihen zu untersuchen, nutzten wir in dieser Arbeit zwei verschiedene Mengen an Zeitreihen. Die Anzahl der Zeitreihen in den Diagrammen belief sich dabei auf vier und sieben. Wir nahmen durch die Vorstudien an, dass diese beiden Mengen an Zeitreihen für zwei verschiedene Zeitreihenanzahlen geeignet waren, da sich keiner der Teilnehmer\*innen negativ zu diesem Aspekt äußerte.

Von der Untersuchung einer dritten Menge an Zeitreihen, wie Javed et al. [JME10] dies handhabten, haben wir aus Zeitgründen abgesehen. Es wurde versucht, die Studie so zu gestalten, dass Ermüdungseffekte im besten Fall vermieden werden, aber gleichzeitig genug Diagramme bearbeitet werden, sodass die Ergebnisse robust und präzise waren. Im Falle der Einbeziehung einer dritten Anzahl von Zeitreihen hätte man mit Ermüdungseffekten rechnen und/oder die Anzahl der Wiederholungen einer Bedingung heruntersetzen müssen. Letzteres würde wohlmöglich eine niedrigere Robustheit und Präzision der Ergebnisse bedeuten.

## **4.10 Zusammenfassung**

In diesem Kapitel wurden die Rahmenbedingungen, das Design und der Ablauf unserer Studie dargelegt. Das Design umfasste drei Visualisierungstechniken, drei verschiedene Aufgaben und zwei Zeitreihenmengen. Die zur Darstellung benötigten Daten generierten wir durch eine Random Walk-Implementation selbst. Aufgrund der Durchführung einer Online-Studie verwendeten die 48 Teilnehmer\*innen ihre eigenen Geräte. Zusätzlich wurden diverse Designentscheidungen begründet, die sich zum Teil aus den Ergebnissen und dem Feedback von den zwei Vorstudien ergaben. Diese Vorstudien werden im folgenden Kapitel beschrieben.

## 5 Vorstudien

Um sowohl Unklarheiten und mögliche Probleme zu identifizieren, als auch die Durchführbarkeit der Studie zu testen, führten wir zwei Vorstudien (auch Pilotstudien genannt) durch. Die erste Vorstudie fand zeitlich vor der zweiten Vorstudie statt, um mögliche größere Probleme frühzeitig identifizieren zu können. Bei beiden Vorstudien absolvierte eine kleine Anzahl von Teilnehmer\*innen die drei verschiedenen Aufgaben der Studie inklusive einer anschließenden kurzen Befragung. Die Teilnehmer\*innen waren dabei deutsche Muttersprachler\*innen, aber waren ebenso mit der englischen Sprache vertraut.

### 5.1 Erste Vorstudie

Der Ablauf der ersten Vorstudie war nahezu identisch zu dem im Kapitel 4.7 erwähnten Ablauf. Ein Unterschied bestand darin, dass in der ersten Vorstudie jede einzelne Bedingung nicht fünf Mal, sondern drei Mal wiederholt wurde. Grund dafür war, dass wir zu diesem Zeitpunkt die Durchführbarkeit der Studie und mögliche technische Probleme identifizieren wollten. Die zeittechnischen Ergebnisse und die Korrektheit der Teilnehmer\*innen hatten dabei eine untergeordnete Rolle, weshalb wir die Teilnehmer\*innen nicht zu lange mit der Vor-Vorstudie beschäftigen wollten. Zudem existierten zum Zeitpunkt der Vor-Vorstudie noch keine szenischen Beschreibungen für die Aufgaben. Aufgrund der örtlichen Entfernung absolvierte ein Teil der Teilnehmer\*innen die Aufgaben über ein Remote-Programm, wodurch gewisse visuelle Abweichungen entstehen konnten. Die restlichen Teilnehmer\*innen absolvierten die Studie nach Möglichkeit an demselben Gerät vor Ort. Den Teilnehmer\*innen war dabei die Möglichkeit gegeben, Fragen zu stellen.

#### 5.1.1 Probleme

Der Großteil der Teilnehmer\*innen der ersten Vorstudie klagte über den überwiegenden Grünanteil der Farben der Zeitreihen. Insbesondere ein ockerfarbenes Braun und ein Orange stellte die Teilnehmer\*innen vor Probleme. Dadurch kam es vor, dass die Teilnehmer\*innen bei der Auswahl des Radiobuttons erneut vergleichen musste, ob es sich bei der Zeitreihe um die ockerfarbene Zeitreihe oder die orangene Zeitreihe handelte. Solche extra Vergleiche, die nicht zu den eigentlichen Aufgaben der Studie zählen, dauern eine gewisse zusätzliche Zeit. Dadurch könnten die gemessenen Zeiten verfälscht werden. Da unter den Klagenden nicht nur Personen lagen, die unter einer Rot-Grün-Schwäche leiden, wurde das ockerfarbene Braun nach Auswertung der Befragungen der Vorstudie durch ein dunkles Grau der genutzten Farbpalette *Dark2* ersetzt.

### **5.1.2 Sonstiges Feedback**

Die meisten Teilnehmer\*innen gaben an, die Aufgaben nach dem Lesen der Aufgabenstellung verstanden zu haben. Falls Unklarheiten aufkamen, wurden diese mit Hilfe der Übungsdurchgänge beseitigt.

## **5.2 Zweite Vorstudie**

Der Ablauf der zweiten Vorstudie war identisch zu dem in Kapitel 4.7 erwähnten Ablauf.

### **5.2.1 Probleme**

Nachdem schon in der ersten Vorstudie Schwierigkeiten mit dem Erkennen der Farben auftraten, äußerten sich erneut einige Teilnehmer\*innen über die Farben. Dabei beinhaltete unsere Farbpalette zwei verschiedene Grüntöne, die besonders bei den vergleichsweise dünnen Linien im Liniendiagramm für Schwierigkeiten sorgten. Daraufhin wurden diese Grüntöne durch zwei andere Grüntöne des ColorBrewers ersetzt. Obwohl es sich bei den neu eingefügten Farben um farbenblind-freundliche Farben handelte, konnte eine verfälschte Wahrnehmung der Farben durch eine Farbsehschwäche nicht ausgeschlossen werden.

Neben den Schwierigkeiten mit den Farben wünschten sich einige Teilnehmer\*innen, dass die Linien der Zeitreihen in den Small Multiples und den Liniendiagrammen dicker gewesen wären. Nach Diskussion dieses Themas entschieden wir uns, die Dicke der Linien von 3px auf 4px zu erhöhen. Diese Liniendicke prüften wir intern und empfanden sie als geeignet.

Eine\*r der Teilnehmer\*innen gab an, Probleme mit der Visualisierungstechnik ThemeRiver gehabt zu haben. Da die Zeitreihen um eine zentrale Achse angeordnet werden, entsteht unterhalb der untersten Zeitreihe bis zur X-Achse eine leere bzw. weiße Fläche. Diese Fläche interpretierte der oder die Teilnehmer\*in während der Übungsdurchgänge als Zeitreihe, wodurch eine Verwirrung seitens des Teilnehmers oder der Teilnehmerin entstand. Da dies zur Technik der Visualisierung gehört und auch sonst von keiner weiteren Person bemängelt wurde, nahmen wir hier keine Änderungen vor.

### **5.2.2 Sonstiges Feedback**

Eine geringe Anzahl der Teilnehmer\*innen gab an, gegen Ende hin öfter auf den Fortschrittsbalken gesehen zu haben. Eine\*r der Teilnehmer\*innen empfand die Anzahl der Stimuli als zu hoch. Die Option, die Anzahl der Wiederholungen jeder Bedingung zu reduzieren, wurde von uns diskutiert und abgelehnt, da eine geringere Anzahl an Wiederholungen zu einem weniger robusten Ergebnis führen könnte. Die fünf Wiederholungen pro Bedingung erachteten wir als angemessen.

Ein paar Teilnehmer\*innen gaben an, dass ihnen die szenischen Beschreibungen zu Beginn einer Aufgabe geholfen haben, da sie dadurch den Nutzen der Aufgabe auf eine mögliche reale Aufgabe ableiten konnten. Zudem gab eine Person an, dass die Szenarien hilfreich gewesen

sein, da sie sich somit in eine Rolle versetzen konnte, wodurch eine zusätzliche Motivation entstand. Jedoch gaben auch ein paar Teilnehmer\*innen an, dass ihnen die Szenarien kaum geholfen hätten. Da dadurch verstärkt der Eindruck entstand, dass das Aufnehmen und Interpretieren einer solchen szenischen Beschreibung subjektiv ist, entschieden wir uns, diese Beschreibungen in der Studie bestehen zu lassen.

### **5.3 Zusammenfassung**

Dieses Kapitel beschrieb die beiden Vorstudien, die zur Identifikation von Problemen und zur Überprüfung der Durchführbarkeit der Studie umgesetzt wurden. Durch die Vorstudien stellten sich diverse Probleme heraus, insbesondere bei der Wahrnehmung von gewissen Farben. Diese Probleme wurden für die Durchführung der Studie diskutiert und nach Möglichkeit behoben. Da die Vorstudien den Zweck hatten, Probleme frühzeitig zu identifizieren, besaßen die Ergebnisse der Antwortzeiten und die Korrektheit hierbei eine untergeordnete Rolle. Die Ergebnisse der richtigen Studie sind jedoch umso relevanter. Diese werden im folgenden Kapitel dargelegt und diskutiert.



## 6 Ergebnisse und Diskussion

Dieses Kapitel enthält die Auswertung der Studie und die Diskussion der jeweiligen Ergebnisse. Dabei werden sowohl die Zeiten und die Korrektheit der Antworten als auch die Beliebtheit der verschiedenen Visualisierungstechniken auf Seiten der Teilnehmer\*innen ausgewertet. Für die Auswertung der Zeiten verwendeten wir Bootstrapping-Konfidenzintervalle, deren Grundidee in Kapitel 2.5.2 erläutert wurde. Das genaue Skript für die Generierung empirischer Bootstrapping-Konfidenzintervalle stammt von Orloff und Bloom [OB14] und ist in ihrer Arbeit zu finden. Die prozentuale Korrektheit der Antworten wurde durch die relative Häufigkeit bestimmt. Am Ende des Kapitels wird ein Blick auf die Resultate verwandter Arbeiten geworfen. Dabei werden die Ergebnisse vergleichbarer Aufgaben und Visualisierungstechniken mit den Ergebnissen der vorliegenden Arbeit verglichen.

Bei der Auswertung durch die Bootstrapping-Konfidenzintervalle ist zu beachten, dass für die Zeiten immer nur die korrekten Antworten beachtet wurden. Die Zeiten der falschen Antworten wurden nicht miteinbezogen. Um pro Teilnehmer\*in nur eine Beobachtung pro Aufgabe zu erhalten, wurde aus den benötigten Zeiten der jeweilige Mittelwert berechnet. Daraus ergaben sich insgesamt 48 Mittelwerte pro Aufgabe. Aus diesen wurde anschließend, wieder pro Aufgabe, der Mittelwert ermittelt. Anhand jedes Mittelwertes der Mittelwerte wurde das Verfahren der Bootstrapping-Konfidenzintervalle mit 2000 Stichprobenwiederholungen angewandt.

Für die Untersuchung der Visualisierungstechniken verlief dies analog. Das bedeutet, dass pro Teilnehmer\*in eine Beobachtung aus der Kombination Aufgabe und Visualisierungstechnik (A x V – z.B. Aufgabe: Maximum; Visualisierungstechnik: Liniendiagramm) entstand. Für die Untersuchung der Zeitreihenanzahl wurde der Kombination noch die jeweilige Zeitreihenanzahl beigelegt (A x V x Z – z.B. Aufgabe: Maximum; Visualisierungstechnik: Liniendiagramm; Zeitreihen: 4). Somit ergab sich pro Bedingung ein Mittelwert, der sich aus den Mittelwerten der benötigten Zeiten zusammensetzte.

### 6.1 Auswertung und Diskussion der Daten

Im Folgenden ist die Auswertung der Daten in drei Ebenen unterteilt. Für das bessere Verständnis ist die Unterteilung der Ebenen in Abbildung 6.1 zu sehen. Auf der ersten Ebene (siehe Kapitel 6.1.1) werden lediglich die Konfidenzintervalle und die Korrektheit der einzelnen Aufgaben dargelegt. Die Konfidenzintervalle und die Korrektheit für die verschiedenen Visualisierungstechniken werden auf Ebene zwei (siehe Kapitel 6.1.2) separiert nach Aufgabe beschrieben. Dabei sind die Daten beider Zeitreihenmengen zusammengefasst. Zusätzlich dazu wird auf dieser Ebene auf die Ergebnisse zwischen Teilnehmer\*innen mit und ohne Rot-Grün-Sehschwäche eingegangen. Ebene drei (siehe Kapitel 6.1.3) betrachtet

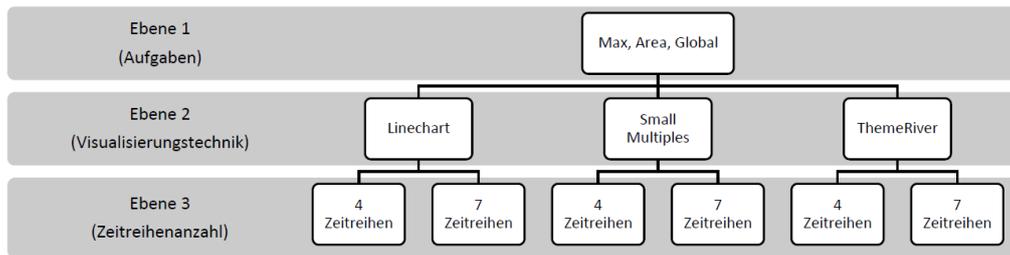


Abbildung 6.1: Darstellung der Ebenen für die Auswertung.

die Konfidenzintervalle und die Korrektheit der Visualisierungstechniken zusätzlich separiert nach Zeitreihenanzahl.

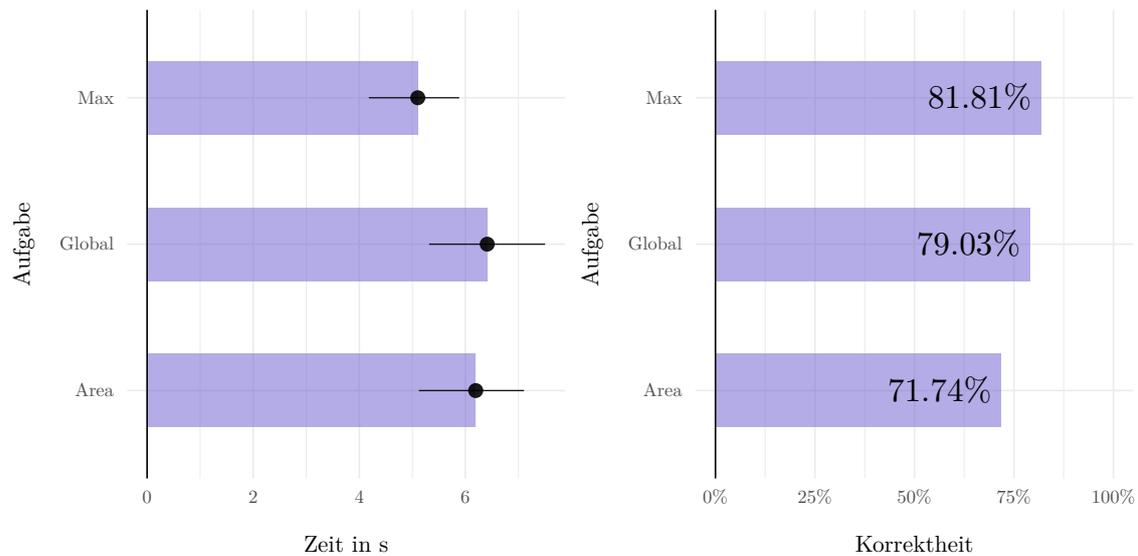
Im Folgenden wird davon gesprochen, dass eine Visualisierungstechnik schneller/langsamer als oder gleich schnell sein kann wie eine der anderen Visualisierungstechniken. Darunter ist die Zeit zu verstehen, die es benötigte, eine Aufgabe anhand eines Stimulus, der durch die jeweilige Visualisierungstechnik dargestellt wird, korrekt zu lösen. Des Weiteren ist davon die Rede, dass sich Konfidenzintervalle ähnlich sind. *Ähnlich* bedeutet hierbei, dass die jeweiligen Konfidenzintervalle größtenteils denselben Wertebereich umfassen.

### 6.1.1 Ebene 1

Hierbei werden die Ergebnisse für jede Aufgabe dargelegt und verglichen. In der Abbildung 6.2 sind die Ergebnisse der jeweiligen Aufgabe zu sehen.

**Diskussion** In Kapitel 4.1 definierten wir die Aufgaben Maximum und Global als einfachere Aufgaben, während wir die Aufgabe Area als komplizierte Aufgabe erachteten. Anhand der Ergebnisse scheint es, dass die Schwierigkeit der Aufgabe nur bedingt einen Einfluss auf die benötigte Antwortzeit hat. Wir erwarteten, dass die Antwortzeit für die Aufgabe Maximum geringer sein würde als die der anderen beiden Aufgaben (siehe E1). Zwar ist es, wie oben genannt, möglich, dass die Aufgabe Maximum eine geringere Antwortzeit benötigt, jedoch ist dies anhand der Konfidenzintervalle bei einer erneuten Stichprobe auch andersherum möglich. Somit kann die Aufgabe Maximum auch länger dauern als eine der anderen beiden Aufgaben. Obwohl bei der Aufgabe Area mehr Informationen extrahiert werden mussten, schnitt diese nicht erkennbar langsamer ab als die beiden einfacheren Aufgaben.

Der unterschiedliche Schwierigkeitsgrad der einzelnen Aufgaben ist jedoch anhand der Korrektheit erkennbar. Bei der Aufgabe Area wurden öfter Fehler gemacht als bei den anderen beiden Aufgaben. Dies ist aufgrund der Menge der Informationen, die extrahiert und verglichen werden müssen, nicht überraschend. Ein möglicher Faktor für das Ergebnis könnte die Unordnung in einem Diagramm und das Betrachten eines Zeitraums sein. Da bei der Aufgabe Maximum nur die Zeitreihe mit dem höchsten Wert interessant ist, ist die Unordnung unterhalb der möglichen höchsten Zeitreihen nicht ausschlaggebend für die Extraktion der Informationen. Zusätzlich wird hierbei nur ein einziger Punkt betrachtet,



(a) Auswertung der benötigten Zeit durch (b) Darstellung der Korrektheit der Teilnehmer\*innen bei den jeweiligen Aufgaben. Eine Darstellung mit Bootstrapping-Konfidenzintervallen.

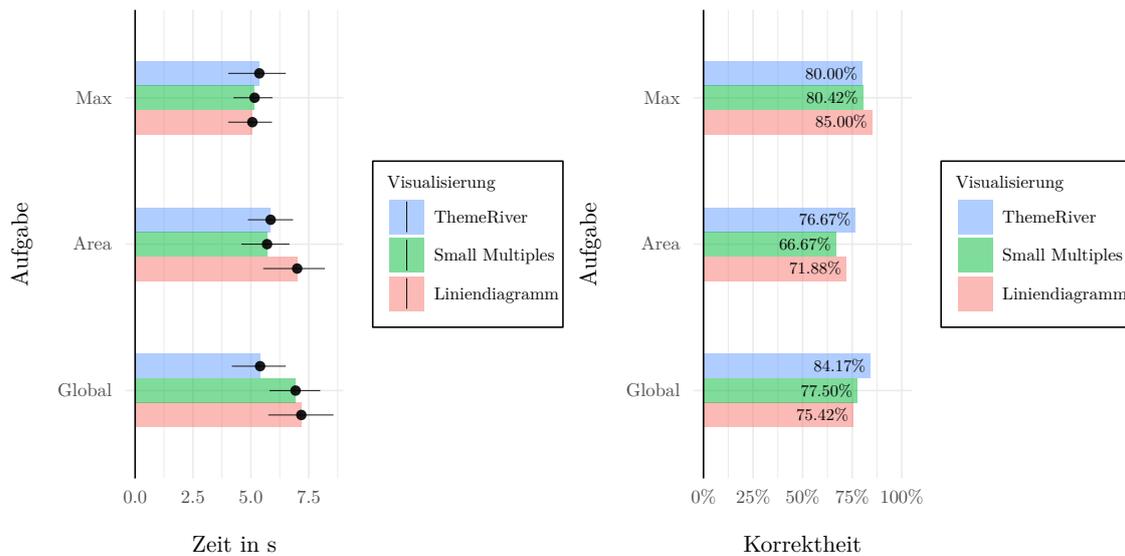
Abbildung 6.2: Darstellung der Auswertung der ersten Ebene.

wodurch die Menge an zu extrahierenden und zu vergleichenden Informationen geringer ausfällt. Ebenso kann bei der Aufgabe Global die Unordnung in einem Diagramm nur bedingt störend sein, da auch hier nur zwei Punkte und kein Zeitraum betrachtet wird. Bei einer noch höheren Anzahl an Zeitreihen ist es jedoch denkbar, dass sich auch hier die Unterscheidung der aggregierten Werte als fehleranfälliger herausstellen kann. Da insbesondere die Aufgabe Global nur zwei Antwortmöglichkeiten umfasste (Punkt A oder Punkt B), war nicht auszuschließen, dass Teilnehmer\*innen versuchten, die Antwort zu erraten.

### 6.1.2 Ebene 2

Hierbei werden die Ergebnisse für jede Visualisierungstechnik nach Aufgabe separiert dargelegt und verglichen. In der Abbildung 6.3 sind die Ergebnisse in Form von paarweisen Konfidenzintervallen der Antwortzeiten und Abbildungen der Korrektheit der jeweiligen Techniken und Aufgaben zu sehen. Im weiteren Verlauf werden diese Ergebnisse diskutiert.

**Diskussion** Ursprünglich erwarteten wir, dass das Liniendiagramm bei der Aufgabe Maximum schneller abschneiden würde als die anderen beiden Visualisierungstechniken (siehe E2). Den Antwortzeiten zufolge spielt dies jedoch keine Rolle. Es ist möglich, dass die Teilnehmer\*innen bereits bei der Betrachtung der Zeitreihen herausfilterten, welche Werte generell als Lösung in Frage kommen. Somit konnte man sich möglichst schnell auf die wichtigen Werte konzentrieren.



(a) Auswertung der benötigten Zeit durch eine Darstellung mit Bootstrapping-Konfidenzintervallen. (b) Darstellung der Korrektheit der Teilnehmer\*innen bei den jeweiligen Aufgaben.

Abbildung 6.3: Darstellung der Auswertung der zweiten Ebene.

Für die Korrektheit der Aufgabe Maximum erwarteten wir, dass das Liniendiagramm besser abschneiden würde als der ThemeRiver und dieser besser als die Small Multiples (siehe E3). Wir fanden heraus, dass das Liniendiagramm ein wenig besser abschneidet als die anderen beiden Visualisierungstechniken. Dies kann daran liegen, dass sich die Betrachter\*innen im Vorhinein schnell auf die höchsten Werte fokussieren und anhand derer einen oder mehrere direkte Vergleiche innerhalb des Diagramms tätigen können. Hierbei liegen die Zeitreihen nah beieinander, wodurch Vergleiche vereinfacht werden, während in den anderen beiden Visualisierungen die Distanz zwischen den Zeitreihen größer ausfällt. Die Korrektheit der Small Multiples war nahezu identisch mit der des ThemeRivers. Dies könnte wie auch bei dem Liniendiagramm daran liegen, dass niedrige Werte im Vorhinein von den Betrachter\*innen ignoriert werden können. Durch dieses Ausschlussverfahren wird die Auswahl eingegrenzt, was in einer geringeren Fehlerwahrscheinlichkeit resultieren kann.

Bei der Aufgabe Area erwarteten wir, dass sowohl die Antwortzeit des ThemeRivers geringer ausfallen würde als die Antwortzeiten der anderen beiden Visualisierungstechniken (siehe E4). Wir fanden heraus, dass sich die Zeiten der Visualisierungstechniken nur geringfügig unterschieden. Das Liniendiagramm kann dabei jedoch ein wenig langsamer sein. Wie oben aber genannt, können der ThemeRiver und die Small Multiples zwar schneller sein als das Liniendiagramm, jedoch ist dies bei einer anderen Stichprobe ebenso andersherum möglich. Ein möglicher Grund für die erhöhte Zeit des Liniendiagramms könnte die angesprochene Unordnung in einem Diagramm sein. Durch diese ist es in einem Liniendiagramm erschwert, einer Zeitreihe über eine gewisse Zeitspanne hinweg fokussiert zu folgen.

Zudem erwarteten wir auch, dass die Korrektheit des ThemeRivers höher ausfallen würde als die der anderen Visualisierungstechniken (siehe E5). Unsere Resultate ergaben, dass der ThemeRiver für diese Aufgabe wie erwartet die höchste Korrektheit aufwies. Auffällig ist hier, dass trotz ähnlicher Antwortzeit die Small Multiples eine deutlich geringere Korrektheit aufweisen als der ThemeRiver. Dies könnte daran liegen, dass bei den Small Multiples die Vergleiche der Zeitreihen über mehrere Diagramme hinweg getätigt werden müssen. Kombiniert mit der Schwierigkeit der Aufgabe könnte dies in der geringeren Korrektheit resultieren. Aufgrund der niedrigen Korrektheit kann nicht ausgeschlossen werden, dass es Teilnehmer\*innen gab, die die Small Multiples bei dieser Aufgabe nicht verstanden haben. Das Liniendiagramm hingegen weist eine höhere Korrektheit auf als die Small Multiples, schneidet jedoch schlechter ab als der ThemeRiver. Grund hierfür könnten die oben genannte Unordnung und die Überlappungen sein.

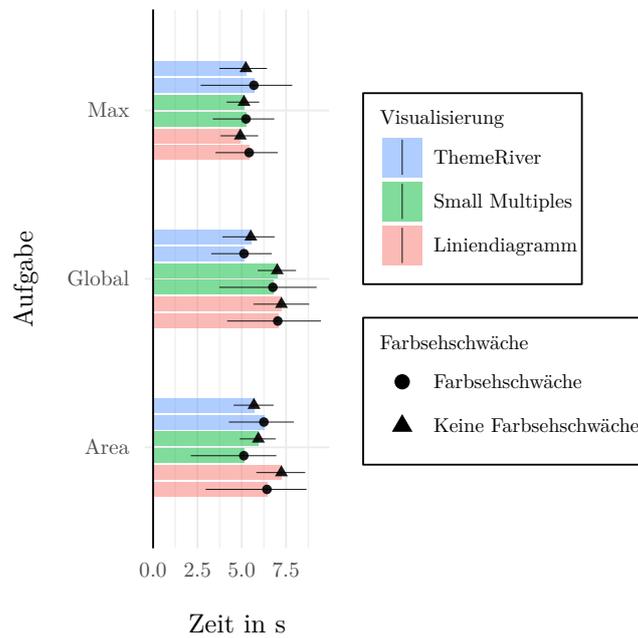
Anfänglich erwarteten wir, dass der ThemeRiver für die Aufgabe Global eine kürzere Antwortzeit benötigen würde als das Liniendiagramm und die Small Multiples (siehe E6). Wir fanden heraus, dass es möglich ist, dass der ThemeRiver schneller abschneidet und es einen Effekt geben kann. Jedoch gibt es kein starkes Argument für den ThemeRiver, da die Daten bei einer erneuten Stichprobe ebenso anders herum ausfallen können. Das bedeutet, dass auch hier die Wahrscheinlichkeit besteht, dass sowohl das Liniendiagramm als auch die Small Multiples bei einer erneuten Stichprobe schneller sein können. Ein\*e Teilnehmer\*in der Vorstudie gab zudem an, er oder sie betrachtete in einem Liniendiagramm meist die Dichte der Zeitreihen an den gegebenen Punkten und fällte anhand dessen eine Entscheidung. Somit spielten für diese\*n Teilnehmer\*in die Werte der Zeitreihen bei dieser Aufgabe nur eine untergeordnete Rolle. Es ist möglich, dass dadurch die Menge an zu extrahierenden Daten abnahm, was in einer geringeren Zeit und einer erhöhten Korrektheit resultierte. Ob dies Teilnehmer\*innen der Hauptstudie so handhabten, ist unklar und wurde nicht weiter untersucht.

Insgesamt scheint es, dass es zwischen den Antwortzeiten der Visualisierungstechniken keine großen Unterschiede gibt. Ebenso liegt die Korrektheit der Techniken in den meisten Fällen nah beieinander. Falls man sich jedoch für eine der Techniken entscheiden müsste, würde sich der ThemeRiver am ehesten eignen. Dies gilt auch dann, wenn man noch wenig bis keine Informationen über die darzustellenden Daten besitzt.

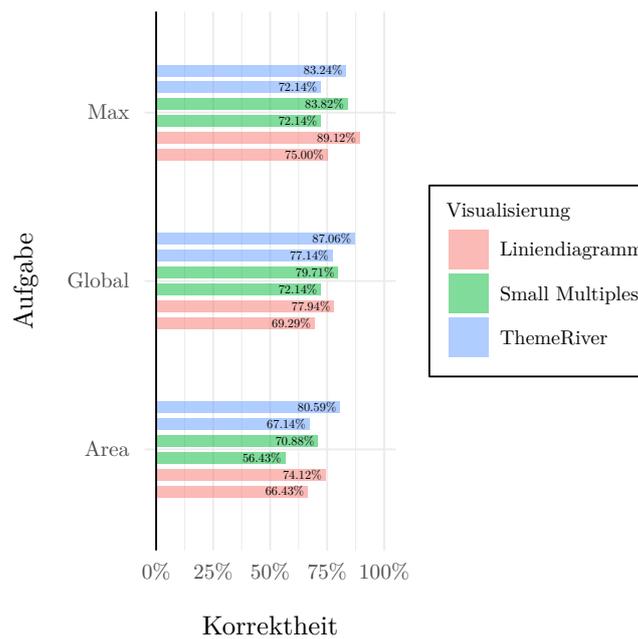
## **Unterscheidung Farbsehschwäche**

Wie sich in den Vorstudien herausstellte, hatten die Teilnehmer\*innen Probleme mit der Farbdarstellung der Zeitreihen. Um zu untersuchen, ob es zwischen den Teilnehmer\*innen mit Rot-Grün-Sehschwäche und den Teilnehmer\*innen ohne Rot-Grün-Sehschwäche einen Unterschied bei der Wahrnehmung der Stimuli gibt, werden diese Daten separiert betrachtet.

In Abbildung 6.4 sind die Ergebnisse in Form von paarweisen Konfidenzintervallen und der Korrektheit zu sehen. In Abbildung 6.5 und 6.6 sind dieselben Ergebnisse nach dem Vorhandensein einer Farbsehschwäche sortiert.



(a) Darstellung der Auswertung der Zeiten der Teilnehmer\*innen mit und ohne Rot-Grün-Farbsehschwäche.



(b) Darstellung der Auswertung der Korrektheit der Teilnehmer\*innen mit und ohne Rot-Grün-Farbsehschwäche. Jeweils der obere Balken verkörpert *Keine Farbsehschwäche*, jeweils der untere Balken *Farbsehschwäche*.

Abbildung 6.4: Darstellung der Auswertung der Ergebnisse der Teilnehmer\*innen mit und ohne Rot-Grün-Farbsehschwäche.

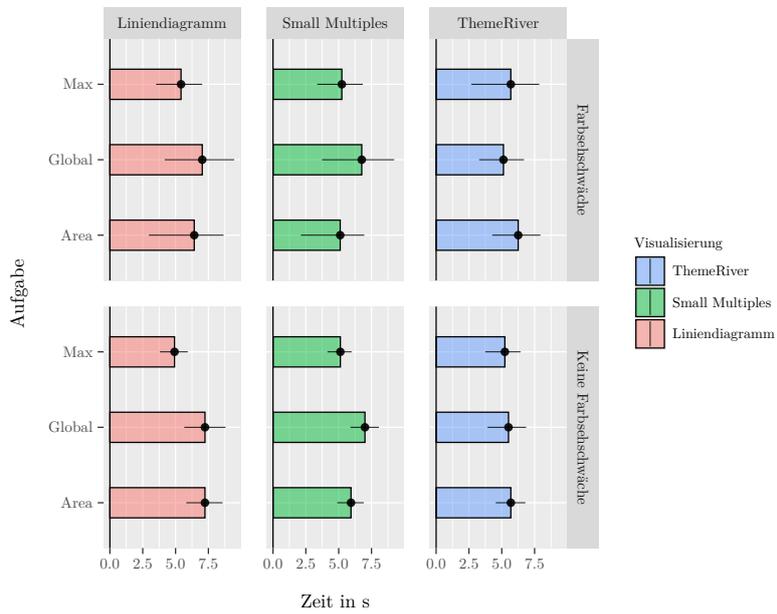


Abbildung 6.5: Darstellung der Auswertung der Zeiten der Teilnehmer\*innen mit und ohne Rot-Grün-Farbsehschwäche.

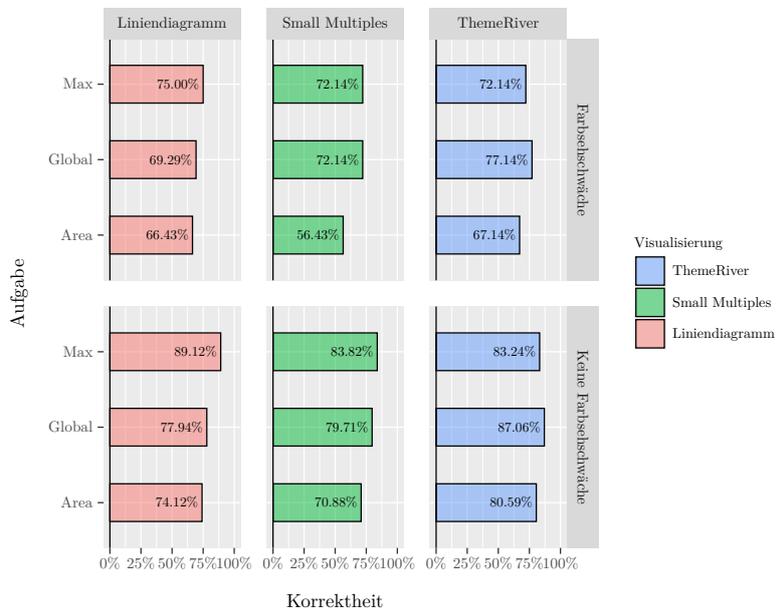


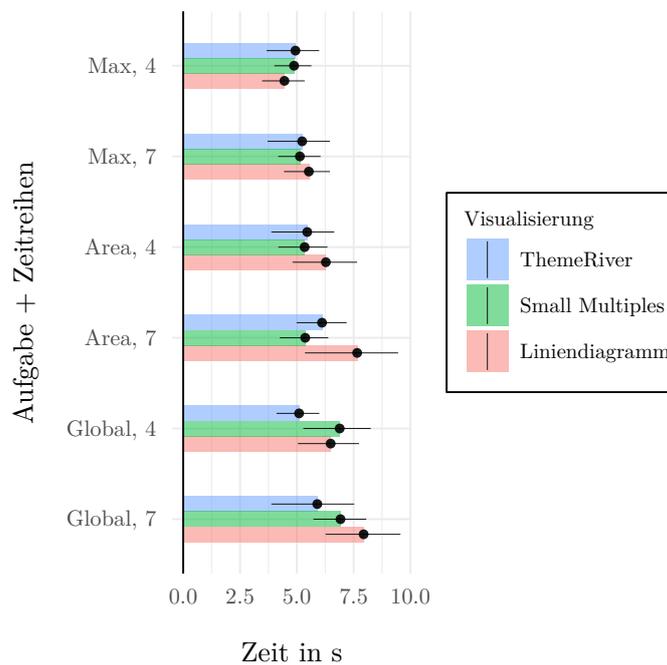
Abbildung 6.6: Darstellung der Auswertung der Korrektheit der Teilnehmer\*innen mit und ohne Rot-Grün-Farbsehschwäche.

**Diskussion** Wir fanden heraus, dass sich die benötigten Antwortzeiten zwischen den Personen mit Farbsehschwäche und ohne Farbsehschwäche weitestgehend kaum bis gar nicht unterschieden. Jedoch ergab sich eine stark verminderte Korrektheit bei den Teilnehmer\*innen mit Farbsehschwäche gegenüber den Teilnehmer\*innen ohne Farbsehschwäche. Ein Grund dafür könnte sein, dass Personen mit einer Farbsehschwäche schlichtweg die Farben in dem Diagramm mit den Farben der Radiobuttons verwechselten und somit eine falsche Antwort auswählten. Überraschend ist dies jedoch bei dem ThemeRiver, da die Antwortmöglichkeiten für die Aufgaben Maximum und Area in der Reihenfolge der geschichteten Zeitreihen sortiert waren. Ein weiterer Grund könnte sein, dass es den Teilnehmer\*innen aufgrund der verfälschten Wahrnehmung die Korrektheit der Antwort egal gewesen ist, da sie die Zeitreihen nur bedingt miteinander vergleichen konnten. Zudem ist überraschend, dass der Unterschied zwischen der Korrektheit der Aufgabe Global höher ausfällt, obwohl die Farbgebung für diese Aufgabe nicht von Relevanz ist. Ein übergreifender Grund könnte sein, dass falsche Antworten bei der Stichprobengröße von 14 Personen stärker ins Gewicht fielen als bei der Anzahl der Teilnehmer\*innen ohne Farbsehschwäche. Außerdem besteht die Möglichkeit, dass Teilnehmer\*innen, die sich wenig bis keine Mühe gegeben haben, als farbsehschwach eingestuft wurden, da sie beim Ishihara-Test eine zufällige Zahl eingaben.

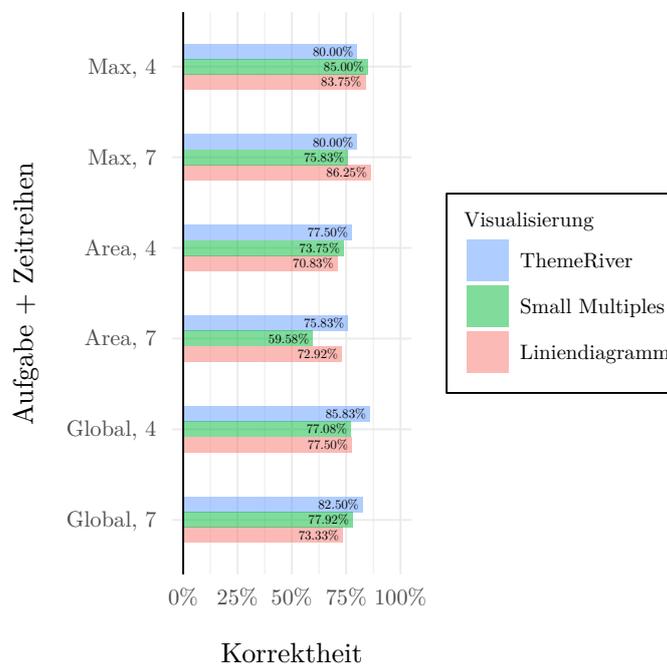
### 6.1.3 Ebene 3

Auf Ebene drei werden die Ergebnisse für jede Visualisierungstechnik nach Aufgabe und Zeitreihenanzahl verglichen. In den Abbildungen 6.7 und 6.8 sind die Ergebnisse in Form von den jeweiligen Konfidenzintervallen und der Korrektheit zu sehen. Diese sind in Abbildung 6.7 zuerst nach den Aufgaben und in Abbildung 6.8 anschließend nach der Zeitreihenanzahl sortiert. In den Abbildungen A.19 und A.20 im Anhang sind dieselben Ergebnisse zu sehen, jedoch sind die Abbildungen zur besseren Übersicht nach der Zeitreihenanzahl aufgeteilt.

**Diskussion** Wir fanden für die Aufgabe Maximum heraus, dass sich die benötigte Zeit mit Erhöhung der Zeitreihenanzahl nicht ausschlaggebend änderte. Das Liniendiagramm kann bei der Darstellung von vier Zeitreihen zwar geringfügig schneller abschneiden, jedoch ist dies bei einer erneuten Stichprobe ebenso andersherum möglich, sodass auch die Darstellung mit sieben Zeitreihen schneller sein kann. Dies kann auch hier daran liegen, dass schon bei der Extraktion der Daten herausgefiltert wird, welche Zeitreihen als mögliche Lösung in Frage kommen. Somit ist es möglich, dass ein gewisser Teil der Zeitreihen ausgeschlossen wird und weniger Vergleiche getätigt werden müssen. Betrachtet man nun dazu die Korrektheit, so ist erkenntlich, dass sich diese für die Shared Space-Techniken kaum ändert. Bei den Small Multiples hingegen nahm die Korrektheit mit der Erhöhung der Zeitreihenanzahl ab. Ein Grund dafür könnte sein, dass der vertikale Platz der einzelnen Diagramme mit Zunahme der Zeitreihenanzahl abnimmt. Dadurch kann es schwieriger sein, Informationen zu extrahieren und anschließend Unterschiede zwischen den Zeitreihenwerten zu erkennen. Eine\*r der Teilnehmer\*innen gab an, dass die Dicke der Linie in den Small Multiples es erschwerte, Werte zu erkennen. Da die Liniendicke von 4px für beide Zeitreihenmenge genutzt wurde, könnte dies ein weiterer Grund für die geringere Korrektheit sein. Insgesamt scheint es jedoch, dass die Anzahl der Zeitreihen für die Extraktion der Daten und das Vergleichen dieser bei einfachen Aufgaben eine untergeordnete Rolle spielt.

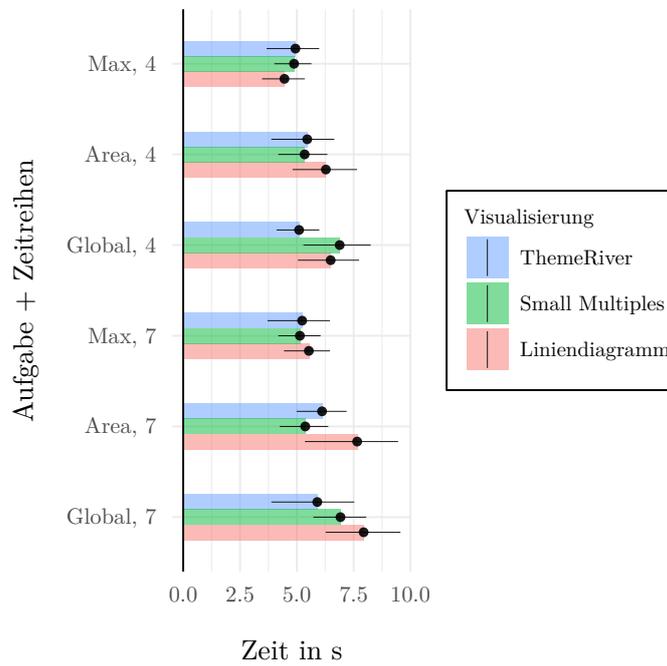


(a) Auswertung der benötigten Zeit durch eine Darstellung mit Bootstrapping-Konfidenzintervallen.

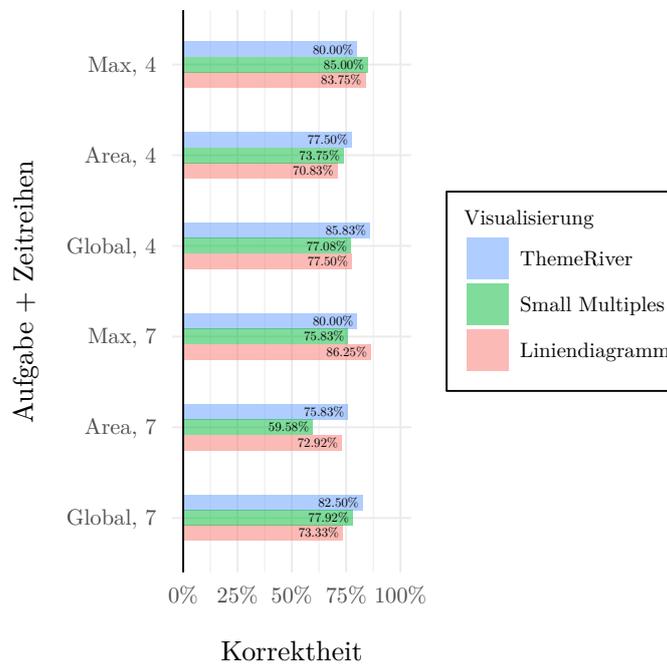


(b) Darstellung der Korrektheit der Teilnehmer\*innen bei den jeweiligen Aufgaben.

Abbildung 6.7: Darstellung der Auswertung der dritten Ebene – gruppiert nach Aufgaben.



(a) Auswertung der benötigten Zeit durch eine Darstellung mit Bootstrapping-Konfidenzintervallen.



(b) Darstellung der Korrektheit der Teilnehmer\*innen bei den jeweiligen Aufgaben.

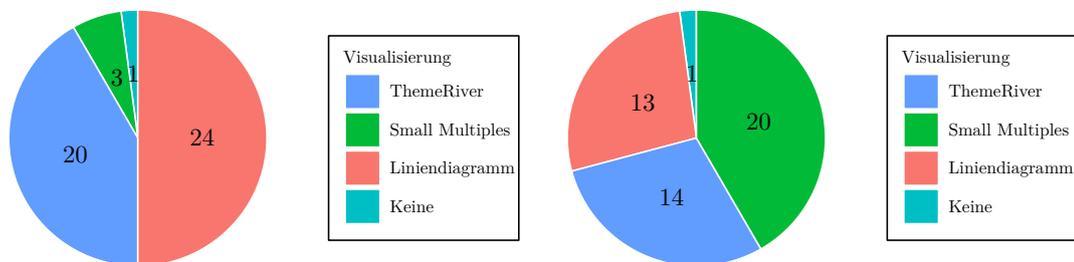
Abbildung 6.8: Darstellung der Auswertung der dritten Ebene – gruppiert nach Zeitreihenanzahl.

Bei der Aufgabe Area erwarteten wir, dass sich Antwortzeit mit Zunahme der Zeitreihenanzahl erhöhen würde (siehe E7). Für die benötigte Antwortzeit machten wir aus, dass sich diese mit Zunahme der Zeitreihenanzahl im Fall des ThemeRivers und der Small Multiples nur geringfügig änderte. Bei dem Liniendiagramm ist es zwar möglich, dass die benötigte Antwortzeit für diese Aufgabe bei einer geringeren Zeitreihenanzahl niedriger ausfällt, jedoch zeigen unsere Ergebnisse auch, dass es bei einer neuen Stichprobe ebenso möglich ist, dass die Darstellung mit einer höheren Anzahl an Zeitreihen schneller sein kann. Ein Grund dafür kann die zunehmende Unordnung und die steigenden Überlappungen der Zeitreihen innerhalb des Liniendiagramms sein. Dadurch ist es möglich, dass die Teilnehmer\*innen eine längere Zeit benötigten, die einzelnen Zeitreihen aufmerksam zu betrachten und anschließend zu vergleichen. Zusätzlich erwarteten wir, dass sich durch die erhöhte Unordnung in den Liniendiagrammen, der geringere vertikale Platz der Small Multiples und die erhöhte Anzahl an Flächen in den ThemeRivern die Korrektheit mit der Zunahme der Zeitreihenanzahl verringern würde (siehe E8). Während die Antwortzeiten der Small Multiples fast identisch blieben, nahm die Korrektheit mit der Erhöhung der Zeitreihenanzahl stark ab. Gründe dafür könnten, wie auch in der Aufgabe Maximum, der geringere vertikale Platz und die Dicke der Linien einer Zeitreihe sein. Es scheint also, dass sich die Erhöhung der Zeitreihen bei komplexeren Aufgaben insbesondere auf die Korrektheit der Small Multiples auswirkt, während lediglich die Antwortzeit des Liniendiagramms höher ausfallen kann.

Für die Aufgabe Global fanden wir heraus, dass sich die Antwortzeiten der Visualisierungstechniken mit Erhöhung der Zeitreihenanzahl nur wenig veränderten. Es könnte bei einer erneuten Stichprobe also möglich sein, dass die Darstellung von sieben Zeitreihen eine kürzere Antwortzeit benötigt. In Bezug auf die Korrektheit sank diese mit der Zunahme der Zeitreihen für den ThemeRiver und das Liniendiagramm geringfügig, während sie für die Small Multiples nahezu gleich blieb. Dies könnte im Falle des Liniendiagramms daran liegen, dass in diesem mehrere Zeitreihen betrachtet und aggregiert werden müssen. Zudem war es möglich, dass die Zunahme der Unordnung die Extraktionen der Werte der Zeitreihen erschwerte. Insgesamt scheint es, dass die Zunahme der Zeitreihenanzahl bei dieser globalen Aufgabe zu geringfügigen Unterschieden zwischen den Antwortzeiten und der Korrektheit führen kann.

## 6.2 Feedback der Teilnehmer\*innen

Die Teilnehmer\*innen wurden nach Beendigung der Aufgaben gebeten, anzugeben, welche der Visualisierungstechniken ihnen am besten und welche am wenigsten gefallen hat (siehe Abbildung 6.9 und 6.10). Ebenso war es möglich, weiteres Feedback abzugeben. Die Hälfte der Teilnehmer\*innen gab an, das Liniendiagramm am meisten gemocht zu haben, wodurch dieses in dieser Studie damit als beliebteste Visualisierungstechnik hervorgeht. Dies kann unter anderem daran liegen, dass das Liniendiagramm heutzutage häufig Anwendung im Alltag findet. Daher sind viele Personen an dieses Diagramm gewöhnt, verstehen es vollumfänglich und können leicht damit umgehen. Der ThemeRiver ist mit nur vier Stimmen weniger als das Liniendiagramm die zweitbeliebteste Visualisierungstechnik der Studie. Bereits in der zweiten Vorstudie gab ein\*e Teilnehmer\*in an, den ThemeRiver als intuitiv wahrgenommen zu haben. Ein Grund für die Beliebtheit dieser Visualisierungstechnik



(a) Beliebtheit der Visualisierungstechniken. (b) Unbeliebtheit der Visualisierungstechniken.

Abbildung 6.9: Darstellung der (Un-)Beliebtheit der Visualisierungstechniken.

könnte die Ästhetik des ThemeRivers sein. Bereits Byron stellte fest, dass die Ästhetik neben der Lesbarkeit eine wichtige Rolle bei der Beliebtheit von Visualisierungen spielt [BW08]. Die Small Multiples erwiesen sich in dieser Studie als unbeliebteste Visualisierungstechnik. Während drei Teilnehmer\*innen angaben, dass die Small Multiples ihnen am besten gefallen haben, gaben 20 Teilnehmer\*innen an, sie am wenigsten gemocht zu haben. Dies kann daran liegen, dass Small Multiples im Alltag eher selten zum Einsatz kommen und somit eventuell für Teilnehmer\*innen neu waren. Es ist daher nicht auszuschließen, dass sich Teilnehmer\*innen mit dieser Visualisierungstechnik überfordert oder unwohl fühlten. Ebenso gilt dieses mögliche Argument auch für den ThemeRiver. Zwar erwies sich dieser als beliebt, dennoch gaben 14 Leute an, diesen am wenigsten gemocht zu haben. Das Liniendiagramm wurde mit 13 Stimmen am seltensten zur unbeliebtesten Visualisierung gewählt. Ein\*e Teilnehmer\*in gab an, keine beliebteste und keine unbeliebteste Visualisierung zu haben.

Neben der Beliebtheit der Visualisierungstechniken gab ein\*e Teilnehmer\*in an, dass manche Farbkontraste der Flächen im ThemeRiver im Vergleich zu benachbarten Zeitreihen die Wahl beeinflussten. Hierbei wurden manche Flächen größer wahrgenommen als bei einem leichteren Kontrast. Dies wurde von uns nicht weiter beobachtet und bedarf weiteren Untersuchungen.

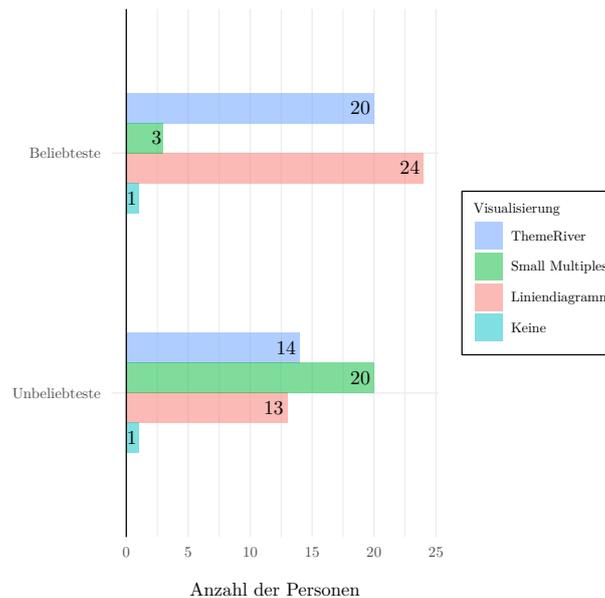


Abbildung 6.10: Alternative Darstellung der (Un-)Beliebtheit der Visualisierungstechniken.

### 6.3 Vergleich von Ergebnissen verwandter Arbeiten

Mit der Aufgabe Maximum führten wir in dieser Arbeit dieselbe wie Javed et al. [JME10] durch. Unsere Ergebnisse für die Antwortzeiten des Liniendiagramms und der Small Multiples unterscheiden sich zu den Ergebnissen der Arbeit von Javed et al. geringfügig. In den Resultaten von Javed et al. ist es möglich, dass die Small Multiples eine höherer Antwortzeit benötigen können als in unserer Arbeit. Die Antwortzeiten für das Liniendiagramm sind jedoch sehr ähnlich zu unseren Messungen. Die Korrektheit der Antworten fällt jedoch in der Studie von Javed et al. höher aus als die in der vorliegenden Arbeit. Zudem verzeichneten Javed et al. in ihrer Arbeit mit der Zunahme der Zeitreihenanzahl höhere Antwortzeiten, während die jeweiligen Antwortzeiten in unserer Arbeit nahezu gleich blieben.

Thudt et al. [TWP+16] führten in ihrer Arbeit dieselbe Aufgabe wie die Aufgabe Global unserer Arbeit durch. Dabei betrachteten sie unter anderem den ThemeRiver. Die Ergebnisse von Thudt et al. ergaben eine deutlich höhere Antwortzeit als unsere Ergebnisse für den ThemeRiver. Während bei Thudt et al. das Konfidenzintervall für die Antwortzeiten des ThemeRivers im Bereich von ungefähr 20 bis 30 Sekunden lagen, befinden sich unser Konfidenzintervall im Bereich von ungefähr 4 bis 6,5 Sekunden. Die Korrektheit der Ergebnisse ist dabei ähnlich zu unseren Messungen.



## 7 Zusammenfassung

Ziel dieser Arbeit war es, die Lesbarkeit von verschiedenen Visualisierungen in Abhängigkeit von der Anzahl der Zeitreihen und der zu absolvierenden Aufgabe zu untersuchen. Um dies zu untersuchen, wurde eine Online-Benutzerstudie mit 48 Teilnehmer\*innen durchgeführt. Diese betrachtete die drei verschiedenen Visualisierungstechniken Liniendiagramm, ThemeRiver und Small Multiples. Zudem umfasste die Studie drei verschiedene Aufgaben, die verschiedene Komplexitätslevel beinhalteten und zwei verschiedene Zeitreihenmengen. Die Lesbarkeit wurde dabei durch die benötigte Zeit zum Lösen einer Aufgabe und der Korrektheit der Antworten gemessen.

Insgesamt scheint es, dass es bei der Wahl der Visualisierungstechnik keine großen Unterschiede zwischen diesen gibt. Falls man sich dennoch (bei noch unbekanntem Umständen/Informationen) für eine der Visualisierungstechniken entscheiden müsste, scheint sich der ThemeRiver besser zu eignen als das Liniendiagramm und die Small Multiples. Der ThemeRiver wies für jede Aufgabenart eine schnelle Antwortzeit auf, während die Korrektheit ebenso hoch gewesen ist. Zudem erwies sich der ThemeRiver bei der Zunahme von Zeitreihen als ebenso robust, was sich in den geringen Unterschieden zwischen den Darstellungen von vier und sieben Zeitreihen widerspiegelt. Die Ästhetik des ThemeRivers hatte neben der Lesbarkeit ebenso Einfluss auf eine hohe Beliebtheit.

Das Liniendiagramm und die Small Multiples scheinen bezüglich den Antwortzeiten ungefähr gleichauf zu liegen. Die Antwortzeiten und Korrektheit hängen bei diesen Visualisierungstechniken von den Bedingungen ab. Das Liniendiagramm erwies sich für einfache Aufgaben (Maximum und Global) ähnlich schnell wie die Small Multiples. Diese können bei der Lösung von komplexeren Aufgaben zwar etwas schneller sein, leiden jedoch dabei unter einer geringeren Korrektheit. Zudem scheint es, dass das Liniendiagramm bei der Zunahme der Zeitreihenanzahl robuster ist als die Small Multiples. Die Korrektheit letzterer nahm mit der Zunahme der Zeitreihenanzahl in zwei von drei Aufgaben erheblich ab, während die Korrektheit des Liniendiagramms vergleichsweise ähnlich blieb. Daher, und aufgrund der Beliebtheit des Liniendiagramms und der Unbeliebtheit der Small Multiples, ergibt sich jedoch eine Tendenz zugunsten des Liniendiagramms.

Zudem zeigte die Studie, dass die Wahl der Farben eine wichtige Rolle bei der Extraktion und dem Vergleich von Daten spielt. Zwar existieren einige Farbpaletten, die sich für die Farbgebung von Visualisierungen als geeignet herausstellten, jedoch ist die Farbwahrnehmung ein subjektiver Prozess. Hinzu kommen diverse weitere Faktoren wie Farbsehschwächen, die sich zusätzlich auf die Extraktion von Daten auswirken.



# A Anhang

Der Anhang beinhaltet die Ergebnisse der Studie, einen beispielhaften Ablauf der Studie und die demographischen Verteilungen der Teilnehmer\*innen. Bei dem beispielhaften Ablauf der Studie wird auf die Darstellungen der Übungsdurchläufe der Aufgaben Area und Global verzichtet, da sie bis auf den Aufgabenumfang deckungsgleich mit dem Übungsdurchlauf der Aufgabe Maximum sind.

## A.1 Ergebnisse

Ebene 1						
Aufgabe	95% Konfidenzintervall		Mittelwert	Korrekte Antworten	Gesamte Antworten	Anteil
	Untere Intervallgrenze	Obere Intervallgrenze				
Max	4,1797	5,8833	5,1020	1178	1440	0,8181
Area	5,1238	7,1044	6,1936	1033	1440	0,7174
Global	5,3169	7,5014	6,4110	1138	1440	0,7903

Tabelle A.1: Tabelle mit den Ergebnissen der ersten Ebene.

Ebene 2							
Aufgabe	Visualisierung	95% Konfidenzintervall		Mittelwert	Korrekte Antworten	Gesamte Antworten	Anteil
		Untere Intervallgrenze	Obere Intervallgrenze				
Max	Liniendiagramm	4,0193	5,9134	5,0655	408	480	0,85
Max	ThemeRiver	4,0243	6,5070	5,3663	384	480	0,8
Max	Small Multiples	4,2455	5,9366	5,1614	386	480	0,8042
Area	Liniendiagramm	5,5331	8,1965	6,9998	345	480	0,7188
Area	ThemeRiver	4,8714	6,8211	5,8536	368	480	0,7667
Area	Small Multiples	4,5887	6,6680	5,6994	320	480	0,6667
Global	Liniendiagramm	5,7539	8,5670	7,1790	362	480	0,7542
Global	ThemeRiver	4,1795	6,5066	5,3994	404	480	0,8417
Global	Small Multiples	5,8034	7,9977	6,9310	372	480	0,775

Tabelle A.2: Tabelle mit den Ergebnissen der zweiten Ebene.

Ebene 3 mit 4 Zeitreihen							
Aufgabe	Visualisierung	95% Konfidenzintervall		Mittelwert	Korrekte Antworten	Gesamte Antworten	Anteil
		Untere Intervallgrenze	Obere Intervallgrenze				
Max	Liniendiagramm	3,4619	5,3264	4,4533	201	240	0,8375
Max	ThemeRiver	3,7412	5,9183	4,9373	192	240	0,8
Max	Small Multiples	4,0417	5,6927	4,8764	204	240	0,85
Area	Liniendiagramm	4,7141	7,6384	6,2780	170	240	0,7083
Area	ThemeRiver	3,8996	6,6465	5,4567	186	240	0,775
Area	Small Multiples	4,1681	6,4111	5,3420	177	240	0,7375
Global	Liniendiagramm	5,1273	7,7277	6,4883	186	240	0,775
Global	ThemeRiver	4,1262	6,0718	5,0991	206	240	0,8583
Global	Small Multiples	5,3107	8,2362	6,8855	185	240	0,7708

Tabelle A.3: Tabelle mit den Ergebnissen der dritten Ebene mit vier Zeitreihen.

Ebene 3 mit 7 Zeitreihen							
Aufgabe	Visualisierung	95% Konfidenzintervall		Mittelwert	Korrekte Antworten	Gesamte Antworten	Anteil
		Untere Intervallgrenze	Obere Intervallgrenze				
Max	Liniendiagramm	4,3937	6,4526	5,5281	207	240	0,8625
Max	ThemeRiver	3,7396	6,4672	5,2325	192	240	0,8
Max	Small Multiples	4,1448	6,0658	5,1332	182	240	0,7583
Area	Liniendiagramm	5,2806	9,3949	7,6539	175	240	0,7292
Area	ThemeRiver	4,9218	7,1206	6,1081	182	240	0,7583
Area	Small Multiples	4,2855	6,3821	5,3689	143	240	0,5958
Global	Liniendiagramm	6,2736	9,6473	7,9388	176	240	0,7333
Global	ThemeRiver	3,8502	7,4952	5,8995	198	240	0,825
Global	Small Multiples	5,6486	8,0616	6,9212	187	240	0,7792

Tabelle A.4: Tabelle mit den Ergebnissen der dritten Ebene mit sieben Zeitreihen

## A.2 Ablauf der Benutzerstudie

### Welcome to this user study!

Please read the following information carefully.

**Important:** For this user study, Google Chrome or Mozilla Firefox is needed. Also, please make sure, you are using a device with a mouse and a keyboard. Do not use mobile devices like smartphones or tablets for this study.

As part of my bachelor thesis, I am investigating different techniques for visualizing temporal multivariate data in diagrams. In this study, you will see three different techniques, including **line charts**, **themerrivers**, and **small multiples**.

Each of these visualization techniques can be applied in different fields of use, for example: with additional relevancy owed to the ongoing corona crisis, in medicine, in finance (stocks, revenues), population, etc., and brings about varying advantages and disadvantages.

Thereupon your job will be completing three diverse tasks with the different visualization types. Before starting, each of these tasks will be explained to you and you have to complete at least one practice run. For a better understanding of the task, you will be notified during the practice run, if your answer was correct or false. Afterward, you can start with the main run or, if you feel unsure about the task, repeat the practice run.

During completion of the tasks, please assume a normal sitting position and avoid coming closer to the screen to get a better view of the diagrams. If you usually use a reading aid, please use it for this study as well. Please **do not** use any external aid (e.g. your fingers, hands, or other pointing devices).

Please note that the time needed for completing the different tasks will be measured. Each task will be measured separately. The measurement starts with the beginning of the main run. During the practice run, you can take as much time as you need. If you need a break, you will be notified of the appropriate moment (usually between the tasks).

Try to complete the tasks **correctly** rather than aiming for a faster completion time.

Before you can proceed to the first task, there will be a demographic questionnaire, a simple resolution calibration task, and a short color vision test in the next steps.

All acquired data and answers will be sent upon completion of the feedback at the end.

Enjoy and have fun!

Abbildung A.1: Darstellung des Einführungsbildschirms.

Please fill in all details about yourself.

Fields marked with an \* are required.

\* Please choose your age group.  <18  18-20  21-30  31-40  41-50  50<

\* Do you need any reading aid?  Yes  No

\* Please select your highest degree or level of education.  No degree (yet)  Less than high school diploma  High School  Bachelor's Degree  Master's Degree  Ph.D. or higher

What job do you have?

[Next](#)

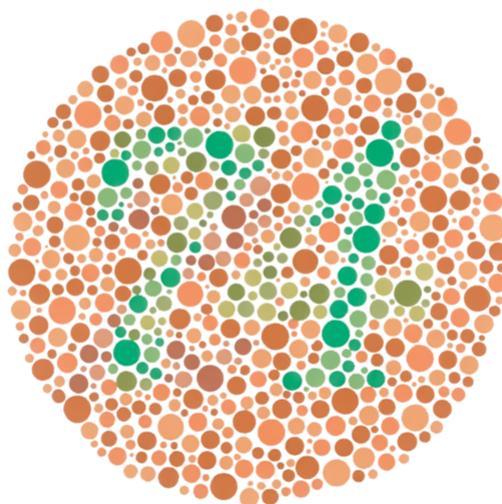
Abbildung A.2: Darstellung des demographischen Fragebogens.

Please take a credit card or a card with similar size and drag the rectangle below to match the size of the card.

[Next](#)

Abbildung A.3: Darstellung der Skalierungsaufgabe für die Anzeige der Visualisierungen.

Please enter the number you see in the image below.



Next

Abbildung A.4: Darstellung der Ishihara-Tafel für den Test einer Rot-Grün-Schwäche.

### Task 1

Assume the following scenario:

You work as a newly hired statistician for a famous movie production company. Your boss is always very busy and wants you to analyze the movie revenue from several movies. He needs to know the movie with the highest revenue from a specific moment.

For this assignment, he provides you the needed diagrams and marked the moment, you have to pay attention to, with an A. Your job is to select the movie with the highest income at the marked spot.

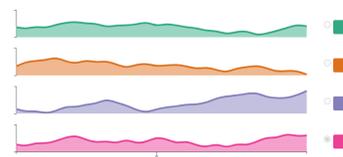
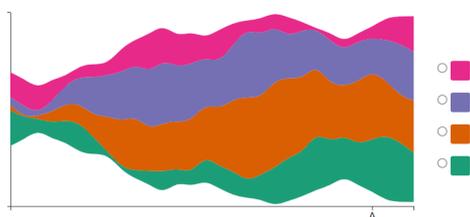
But before you can start, your boss is skeptical, if you ever analyzed these types of diagrams. That's why he gave you a few diagrams for practicing the assignment.

At this task, you will see the different mentioned diagrams with **one** marked spot A in it.

Please indicate the time series with the **highest value** at the marked spot.

Start practice run

Abbildung A.5: Darstellung der szenischen Beschreibung der ersten Aufgabe vor dem Übungsdurchlauf.



Correct answer

- (a) Beispielhafte Darstellung eines ThemeRivers mit vier Zeitreihen bei der Aufgabe Maximum.
- (b) Beispielhafte Darstellung einer korrekten Antwort während des Übungsdurchlaufs bei der Aufgabe Maximum.

Abbildung A.6: Zwei Beispiele des Übungsdurchlaufs für die Aufgabe Maximum.

### Task 1

Assume the following scenario:

You work as a newly hired statistician for a famous movie production company. Your boss is always very busy and wants you to analyze the movie revenue from several movies. He needs to know the movie with the highest revenue from a specific moment.

For this assignment, he provides you the needed diagrams and marked the moment, you have to pay attention to, with an A. Your job is to select the movie with the highest income at the marked spot.

At this task, you will see the different mentioned diagrams with **one** marked spot A in it.

Please indicate the time series with the **highest value** at the marked spot.

Start task

Repeat practice run

Abbildung A.7: Darstellung der szenischen Beschreibung der ersten Aufgabe nach dem Übungsdurchlauf.

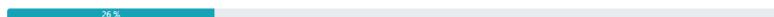
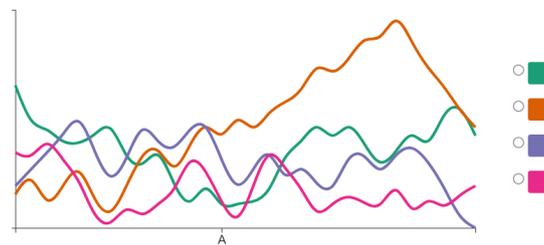


Abbildung A.8: Beispielhafte Darstellung eines Liniendiagramms bei der Aufgabe Maximum inklusive eines Fortschrittsbalkens.

### Task 2

Assume the following scenario:

You own a big restaurant chain with many restaurants in different cities. Now you want to build more restaurants in the city with the most visitors. To do this, you have to know in which city most people have visited your restaurants. For multiple samples, you selected different periods of time to be examined. Originally you wanted your finance employee to do that, but unfortunately, he quit his job a few days ago. So it falls to you to analyze the statistics.

Luckily your former finance employee already created some diagrams and marked the period to be examined with an A and a B. It is now your job to select the city with the most visitors during the marked period.

Since you have less experience with analyzing these diagrams, you practice with a few exercise diagrams first.

At this task, you will see the different mentioned diagrams with **two** marked spots A and B in it.

Please indicate the time series with the **largest area** between the marked spots.

**If you need a break, now is a good time for taking it.**

Start practice run

### Task 2

Assume the following scenario:

You own a big restaurant chain with many restaurants in different cities. Now you want to build more restaurants in the city with the most visitors. To do this, you have to know in which city most people have visited your restaurants. For multiple samples, you selected different periods of time to be examined. Originally you wanted your finance employee to do that, but unfortunately, he quit his job a few days ago. So it falls to you to analyze the statistics.

Luckily your former finance employee already created some diagrams and marked the period to be examined with an A and a B. It is now your job to select the city with the most visitors during the marked period.

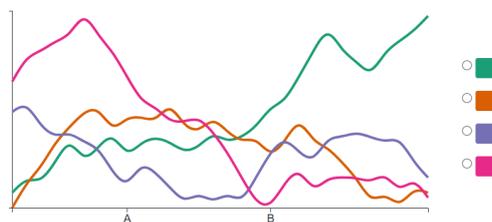
At this task, you will see the different mentioned diagrams with **two** marked spots A and B in it.

Please indicate the time series with the **largest area** between the marked spots.

Start task Repeat practice run

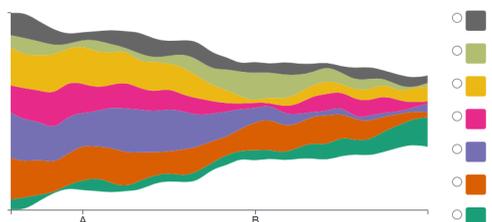
- (a) Szenische Beschreibung der zweiten Aufgabe vor dem Übungsdurchlauf. (b) Szenische Beschreibung der zweiten Aufgabe nach dem Übungsdurchlauf.

Abbildung A.9: Darstellung der szenischen Beschreibung der zweiten Aufgabe.



47%

- (a) Beispielhafte Darstellung eines Liniendiagramms mit vier Zeitreihen bei der Aufgabe Area.



58%

- (b) Beispielhafte Darstellung eines ThemeRivers mit sieben Zeitreihen bei der Aufgabe Area.

Abbildung A.10: Zwei Beispielhafte Darstellungen der Aufgabe Area.

### Task 3

Assume the following scenario:

The World Health Organization (WHO) is looking for the best statisticians and contacted you with the request to analyze data based on the ongoing Coronavirus disease situation. They sent you a bunch of diagrams, which show the number of Corona cases in different countries. In the diagrams, two spots are marked with an A and a B.

The WHO now wants to know, if the Corona cases are increasing or decreasing. To find out, you have to summarize the Corona cases in the various countries at the marked times. You should then decide, if the number of Corona cases is higher at A or at B.

To be sure that you are among the best statisticians, the WHO wants you to analyze a few exercise diagrams first.

At this task, you will see the different mentioned diagrams with **two** marked spots A and B in it.

Please indicate the spot with the **higher sum** of the time series values.

**If you need a break, now is a good time for taking it.**

Start practice run

### Task 3

Assume the following scenario:

The World Health Organization (WHO) is looking for the best statisticians and contacted you with the request to analyze data based on the ongoing Coronavirus disease situation. They sent you a bunch of diagrams, which show the number of Corona cases in different countries. In the diagrams, two spots are marked with an A and a B.

The WHO now wants to know, if the Corona cases are increasing or decreasing. To find out, you have to summarize the Corona cases in the various countries at the marked times. You should then decide, if the number of Corona cases is higher at A or at B.

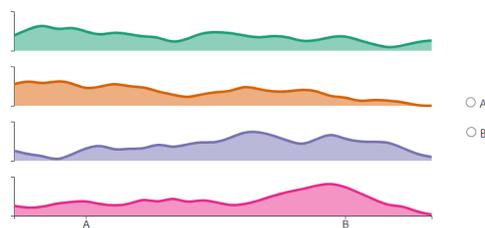
At this task, you will see the different mentioned diagrams with **two** marked spots A and B in it.

Please indicate the spot with the **higher sum** of the time series values.

Start task Repeat practice run

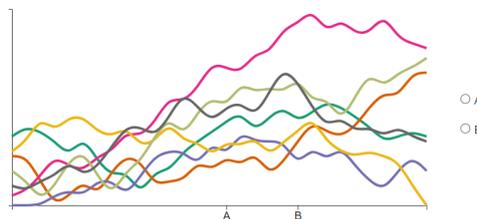
- (a) Szenische Beschreibung der dritten Aufgabe vor dem Übungsdurchlauf. (b) Szenische Beschreibung der dritten Aufgabe nach dem Übungsdurchlauf.

Abbildung A.11: Darstellung der szenischen Beschreibung der dritten Aufgabe.



73%

- (a) Beispielhafte Darstellung von Small Multiples mit vier Zeitreihen bei der Aufgabe Global.



87%

- (b) Beispielhafte Darstellung eines Liniendiagramms mit sieben Zeitreihen bei der Aufgabe Global.

Abbildung A.12: Zwei Beispielhafte Darstellungen der Aufgabe Global.



### A.3 Demographische Daten

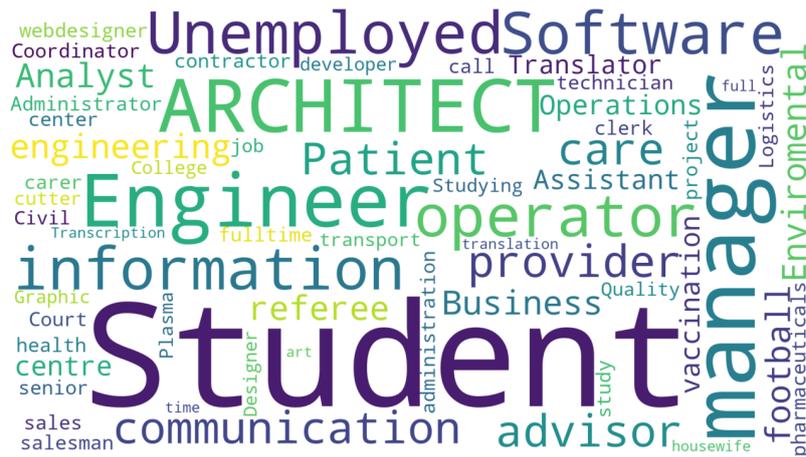


Abbildung A.15: Wordcloud der Berufe der Teilnehmer\*innen.

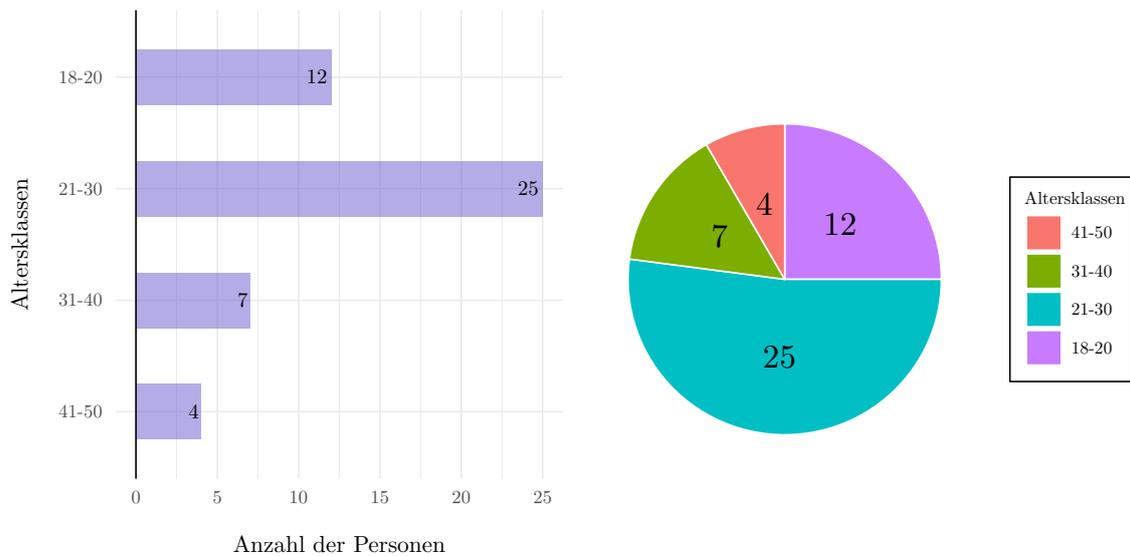


Abbildung A.16: Darstellungen der Verteilung des Alters der Teilnehmer\*innen.

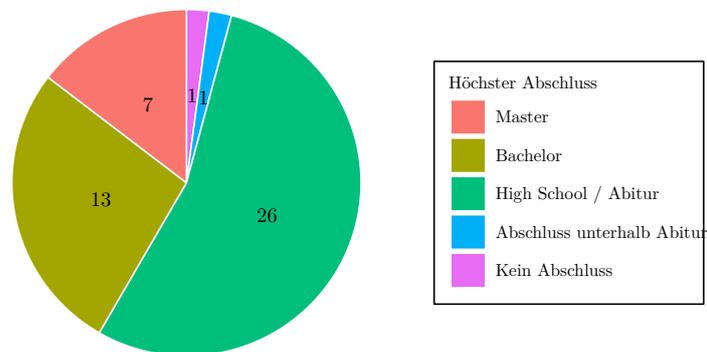
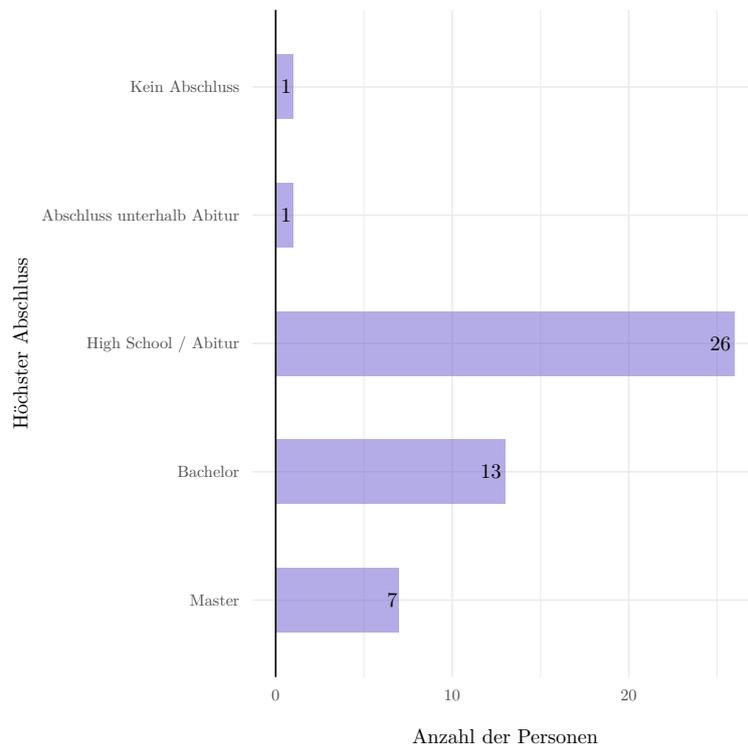
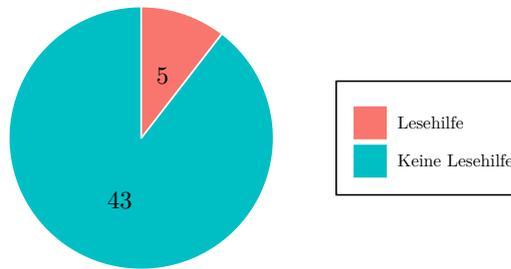
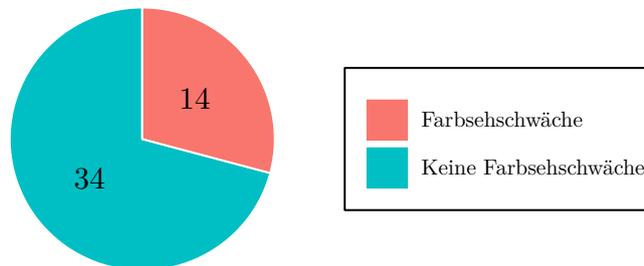


Abbildung A.17: Darstellungen der Verteilung des höchsten Abschlusses der Teilnehmer\*innen.



(a) Verteilung Brillenträger

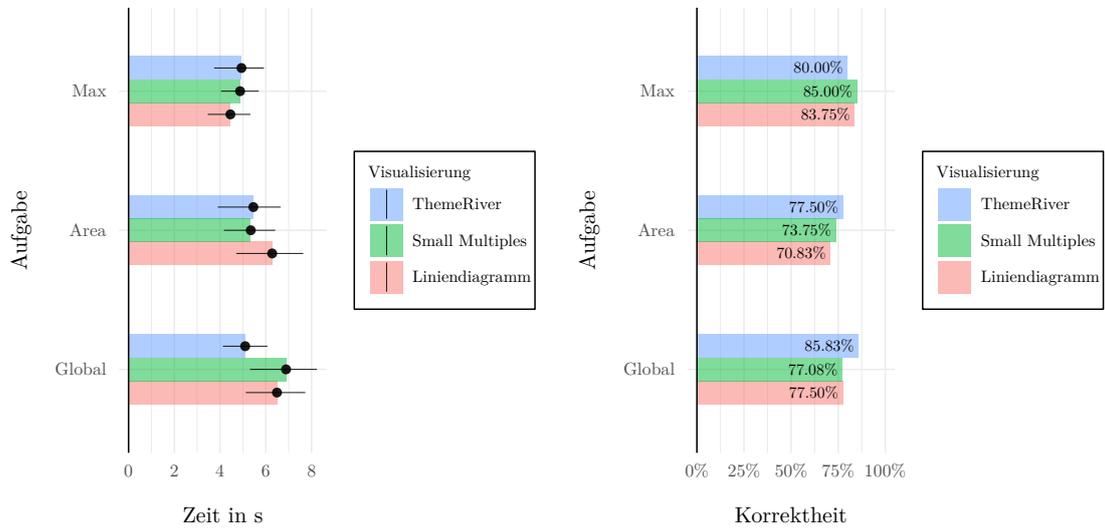


(b) Verteilung Farbsehschwäche

Abbildung A.18: Darstellungen der Verteilung der Teilnehmer\*innen mit und ohne Brille und Farbsehschwäche.

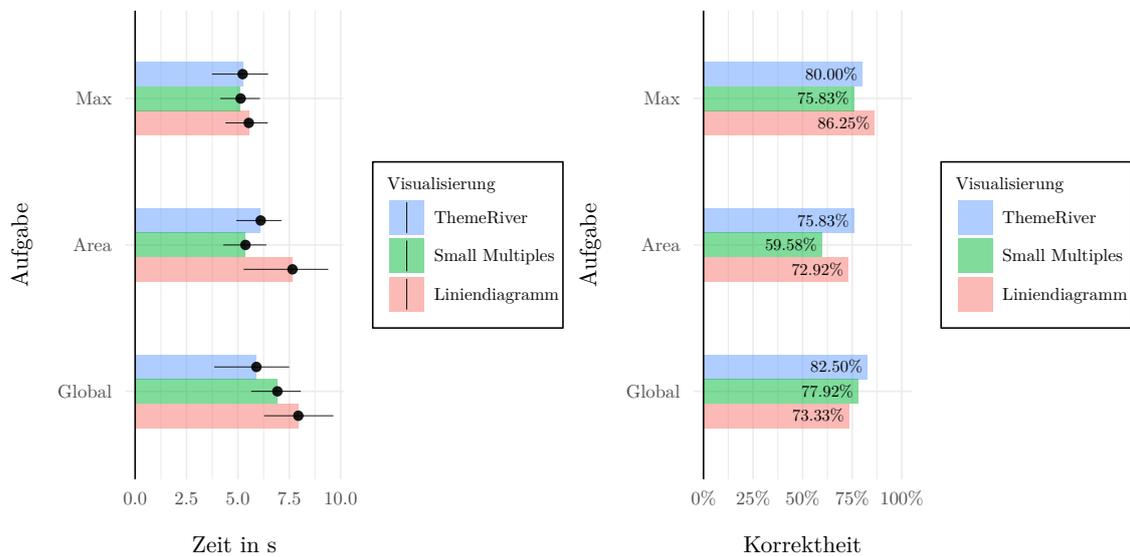
## A.4 Sonstiges

### A.4.1 Auswertungen



(a) Auswertung Ebene 3 Zeit vier Zeitreihen. (b) Auswertung Ebene 3 Korrektheit vier Zeitreihen.

Abbildung A.19: Auswertung der dritten Ebene mit vier Zeitreihen.



(a) Auswertung Ebene 3 Zeit sieben Zeitreihen. (b) Auswertung Ebene 3 Korrektheit sieben Zeitreihen.

Abbildung A.20: Auswertung der dritten Ebene mit sieben Zeitreihen.

## A.4.2 Farbpalette



Abbildung A.21: Farbpalette für vier und sieben Zeitreihen.



# Literaturverzeichnis

- [BBCC08] M. Bloch, L. Byron, S. Carter und A. Cox. *The Ebb and Flow of Movies: Box Office Receipts 1986 — 2008*. Feb. 2008 (zitiert auf S. 8, 25).
- [BVB+13] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva und H. Pfister. „What Makes a Visualization Memorable?“ In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), S. 2306–2315 (zitiert auf S. 3).
- [BW08] L. Byron und M. Wattenberg. „Stacked Graphs – Geometry Aesthetics“. In: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008), S. 1245–1252 (zitiert auf S. 6, 8, 46).
- [BZW+21] C. Bu, Q. Zhang, Q. Wang, J. Zhang, M. Sedlmair, O. Deussen und Y. Wang. „SineStream: Improving the Readability of Streamgraphs by Minimizing Sine Illusion Effects“. In: *IEEE Transactions on Visualization and Computer Graphics* 27.2 (2021), S. 1634–1643 (zitiert auf S. 7, 8).
- [CA15] K. L. Caballero Barajas und R. Akella. „Dynamically Modeling Patient’s Health State from Electronic Medical Records: A Time Series Approach“. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’15. Sydney, NSW, Australia: Association for Computing Machinery, 2015, S. 69–78. ISBN: 9781450336642 (zitiert auf S. 4).
- [CM84] W. S. Cleveland und R. McGill. „Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods“. In: *Journal of the American Statistical Association* 79.387 (1984), S. 531–554 (zitiert auf S. 6–9).
- [Däß99] R. Däßler. „Informationsvisualisierung: Stand, Kritik und Perspektiven“. In: *Methoden/Strategien der Visualisierung in Medien, Wissenschaft und Kunst*. Wissenschaftlicher Verlag Trier, 1999. URL: <https://www.infodata-edepot.de/volltext/fhpotsdam/03021.pdf> (zitiert auf S. 3).
- [Dra16] P. Dragicevic. „Fair statistical communication in HCI“. In: *Modern statistical methods for HCI*. Springer, 2016, S. 291–330 (zitiert auf S. 11–14).
- [FvWSN08] J.-D. Fekete, J. van Wijk, J. Stasko und C. North. „The Value of Information Visualization“. In: Bd. 4950. Juli 2008, S. 1–18. ISBN: 978-3-540-70955-8 (zitiert auf S. 3).

- [Gra] B. Gray. *Box Office Mojo*. [Zuletzt besucht: 05.12.2020]. URL: <https://www.boxofficemojo.com/> (zitiert auf S. 25).
- [HB] M. Harrower und C. A. Brewer. *ColorBrewer 2.0*, [Zuletzt besucht: 11.01.2020]. URL: <https://colorbrewer2.org/> (zitiert auf S. 23, 28).
- [HB03] M. Harrower und C. A. Brewer. „ColorBrewer. org: an online tool for selecting colour schemes for maps“. In: *The Cartographic Journal* 40.1 (2003), S. 27–37 (zitiert auf S. 28).
- [HKA09] J. Heer, N. Kong und M. Agrawala. „Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations“. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2009, S. 1303–1312 (zitiert auf S. 6, 7, 15, 19, 29).
- [JME10] W. Javed, B. McDonnell und N. Elmqvist. „Graphical Perception of Multiple Time Series“. In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), S. 927–934 (zitiert auf S. 5–7, 15, 16, 19, 28, 29, 47).
- [KMSZ06] D. A. Keim, F. Mansmann, J. Schneidewind und H. Ziegler. „Challenges in Visual Data Analysis“. In: *Proceedings of the Conference on Information Visualization. IV '06*. USA: IEEE Computer Society, 2006, S. 9–16. ISBN: 0769526020 (zitiert auf S. 3).
- [Koh13] S. Kohl. *Achromatopsia - Orphanet encyclopedia*. [Zuletzt besucht: 07.02.2021]. 2013. URL: [https://www.orpha.net/consor/cgi-bin/OC\\_Exp.php?lng=EN&Expert=49382](https://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=EN&Expert=49382) (zitiert auf S. 28).
- [Kos07] R. Kosara. „Visualization Criticism - The Missing Link Between Information Visualization and Art“. In: Aug. 2007, S. 631–636. ISBN: 0-7695-2900-3 (zitiert auf S. 3).
- [LI19] C.-Y. Lim und J. In. „Randomization in clinical studies“. In: *Korean journal of anesthesiology* 72.3 (2019), S. 221 (zitiert auf S. 10).
- [LKAV05] J. Lin, E. Keogh, Ada Fu und H. Van Herle. „Approximations to magic: finding unusual medical time series“. In: *18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)*. 2005, S. 329–334 (zitiert auf S. 4).
- [Man10] L. Manovich. „What is visualization“. In: *Lev Manovich: cultural analytics, software studies, new media, digital humanities* (2010) (zitiert auf S. 3).
- [Nat15] National Eye Institute (NEI). *Facts About Color Blindness*. [Zuletzt besucht: 05.12.2020]. Feb. 2015. URL: [https://web.archive.org/web/20160728003639/https://nei.nih.gov/health/color\\_blindness/facts\\_about](https://web.archive.org/web/20160728003639/https://nei.nih.gov/health/color_blindness/facts_about) (zitiert auf S. 9).
- [OB14] J. Orloff und J. Bloom. „Bootstrap confidence intervals“. In: *University Lecture* (2014) (zitiert auf S. 12, 13, 35).

- [Orp07] Orphanet. *Tritanopia - Orphanet encyclopedia*. [Zuletzt besucht: 07.02.2021]. 2007. URL: [https://www.orpha.net/consor/cgi-bin/OC\\_Exp.php?lng=EN&Expert=88629](https://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=EN&Expert=88629) (zitiert auf S. 28).
- [Per17] C. Pernet. „Null hypothesis significance testing: A guide to commonly misunderstood concepts and recommendations for good practice“. In: *F1000Research* 4 (Sep. 2017), S. 621 (zitiert auf S. 13).
- [Pro14] Prolific. *Prolific: Demographics*. <https://app.prolific.co/demographics>. 2014 (zitiert auf S. 23).
- [Spe06] I. Spence. „William Playfair and the Psychology of Graphs“. In: Aug. 2006 (zitiert auf S. 4).
- [SW06] I. Spence und H. Wainer. „Introduction to Playfair’s Commercial and Political Atlas and Statistical Breviary“. In: Jan. 2006, S. 1–35. ISBN: 9780521855549 (zitiert auf S. 4).
- [Tsa05] R. S. Tsay. *Analysis of financial time series*. Bd. 543. John Wiley & Sons, 2005 (zitiert auf S. 4).
- [TWP+16] A. Thudt, J. Walny, C. Perin, F. Rajabiyazdi, L. MacDonald, D. Vardeleon, S. Greenberg und S. Carpendale. „Assessing the Readability of Stacked Graphs“. In: 2016 (zitiert auf S. 8, 15, 16, 19, 21, 24, 25, 47).
- [vWij06] J. J. van Wijk. „Views on Visualization“. In: *IEEE Transactions on Visualization and Computer Graphics* 12.4 (2006), S. 421–432 (zitiert auf S. 4).
- [Won11] B. Wong. *Points of view: Color blindness*. 2011 (zitiert auf S. 9).

Alle URLs wurden zuletzt am 09.02.2021 geprüft.



## **Erklärung**

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Exemplaren überein.

---

Ort, Datum, Unterschrift