

Institut für Visualisierung und Interaktive Systeme

Universität Stuttgart  
Universitätsstraße 38  
D-70569 Stuttgart

Bachelorarbeit

# **Evaluierung von partiellen Kanten in parallelen Koordinaten**

Delani Thiyagarajah

<b>Studiengang:</b>	Informatik
<b>Prüfer/in:</b>	Prof. Dr. Daniel Weiskopf
<b>Betreuer/in:</b>	Dipl.-Inf. Rudolf Netzel
<b>Beginn am:</b>	29. März 2017
<b>Beendet am:</b>	29. September 2017
<b>CR-Nummer:</b>	G.3; I.3.3; I.4.8



## Kurzfassung

Die Visualisierung hochdimensionaler Daten kann auf verschiedene Art realisiert werden. Verwendet wird hierfür häufig eine Streudiagramm-Matrix, welche sich aus allen möglichen Kombinationen von 2D Streudiagrammen zusammensetzt. Mit einer steigenden Anzahl an Dimensionen nimmt der benötigte Platz jedoch quadratisch zu. Eine Alternative, bei der der Platzbedarf linear ansteigt, stellen parallele Koordinaten dar. Die vielen Linienübergänge zwischen verwandten Variablen können in den Parallelen Koordinaten jedoch zu einer visuellen Störung (visual clutter) führen. Um die visuelle Unordnung zu reduzieren, betrachten wir partielle Kanten. Hier stellt sich die Frage, ob solche Diagramme noch lesbar, verständlich und interpretierbar sind. Daher werden in dieser Arbeit die Auswirkungen von partiellen Kanten in parallelen Koordinaten untersucht. Eine Benutzerstudie wird mit 36 Teilnehmern, unter Berücksichtigung der Distanz zwischen den Achsen, der Länge der partiellen Kanten sowie einem Distraktor durchgeführt. Die Resultate zeigen, dass es keine signifikanten Unterschiede bei der Bearbeitungszeit, aber bei der Genauigkeit in Bezug auf die Faktoren Linklänge, Gapfilling und Distanz geben.



## **Abstract**

The visualization of high-dimensional data can be realized in various ways. A scatterplot matrix is often used for this purpose, which is composed of all possible combinations of 2D scatterplots. However, as the number of dimensions increases, the required space increases quadratically. An alternative are parallel coordinates, with which the space requirement increases linearly.

On the downside multiple edge crossings between related variables may lead to visual clutter. In order to lessen this visual confusion, we are going to consider partial edges. This begs the question if such diagrams are still readable, comprehensible and interpretable.

So in this paper we will measure the effects of partial edges in parallel coordinates. A user study with 36 participants will be performed, which takes into account the distance between the axes, the length of the partial edges, and a distractor. The results show, that handling times don't differ significantly, but they revealed an effect on accuracy regarding distance, length and gap filling.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>11</b>
<b>2</b>	<b>Grundlagen</b>	<b>13</b>
2.1	Multivariate Datenanalyse . . . . .	13
2.2	Visualisierungsarten für multivariate Daten . . . . .	13
2.3	Partielle Kanten . . . . .	18
2.4	Korrelation . . . . .	19
<b>3</b>	<b>Verwandte Arbeiten</b>	<b>23</b>
3.1	Vergleich Streudiagramme und Parallele Koordinaten . . . . .	23
3.2	Partielle Kanten in gerichteten Graphen . . . . .	23
3.3	Partielle Kanten in gerichteten Graphen . . . . .	24
3.4	Übersicht Paralleler Koordinaten . . . . .	24
<b>4</b>	<b>Studiendurchführung</b>	<b>27</b>
4.1	Studiendesign und Parameter . . . . .	27
4.2	Pilotenstudie . . . . .	28
4.3	Hypothese . . . . .	29
4.4	Stimuli . . . . .	30
4.5	Rahmenwerk . . . . .	31
4.6	Technische Daten . . . . .	31
4.7	Teilnehmer . . . . .	31
4.8	Studienablauf . . . . .	32
<b>5</b>	<b>Auswertung</b>	<b>39</b>
5.1	Distanz . . . . .	39
5.2	Länge . . . . .	40
5.3	Gap . . . . .	41
5.4	Kombinationen zwischen den drei Faktoren . . . . .	42
<b>6</b>	<b>Zusammenfassung und Ausblick</b>	<b>51</b>
	<b>Literaturverzeichnis</b>	<b>53</b>



# Abbildungsverzeichnis

2.1	Aufbau von parallelen Koordinaten mit fünf Dimensionen . . . . .	14
2.2	Bremsweg in Abhängigkeit der Ausgangsgeschwindigkeit: Die horizontale Achse stellt die Geschwindigkeit und die vertikale Achse den Bremsweg in einem Streudiagramm dar. . . . .	15
2.3	Streudiagramm für die drei Variablen Gewicht, Hubraum und Benzinverbrauch	16
2.4	Beispiel: Mehrdimensional . . . . .	17
2.5	Zwölf verschiedene Chernoff-Gesichter . . . . .	18
2.6	Beispiel: Star-Plots . . . . .	18
2.7	Beispiel: Partielle Kanten in Parallelen Koordinaten . . . . .	19
2.8	Drei Extremwerte der Korrelation für die Visualisierungsarten (A)Streudiagrammen und (B)Parallele Koordinaten . . . . .	20
2.9	Vergleich Streudiagramm und Parallele Koordinaten . . . . .	21
3.1	Darstellung multivariater Daten: (a) Streudiagramm Matrix und (b) Parallele Koordinaten . . . . .	24
3.2	Benutzeroberfläche des Programmtools . . . . .	25
4.1	Genauigkeit in Abhängigkeit der Linklänge: Die horizontale x- Achse stellt die relative Linklänge und die vertikale y-Achse die Genauigkeit in Prozent dar. . . . .	29
4.2	Zeit in Abhängigkeit der Linklänge: Die horizontale x- Achse stellt die relative Linklänge und die vertikale y-Achse die Zeit in Sekunden dar. . . . .	30
4.3	Distanz-Genauigkeit: Genauigkeit in Abhängigkeit der Distanz: Die horizontale x-Achse stellt die Distanz zwischen den Achsen in Pixeln dar und die vertikale y-Achse die Genauigkeit in Prozent. . . . .	31
4.4	Distanz-Zeit: Dauer bis die Korrelation von den Teilnehmern bestimmt wurde in Abhängigkeit der Distanz. Die horizontale Achse ist die Distanz zwischen den Achsen in Pixeln und die vertikale Achse die Zeit in Sekunden . . . . .	32
4.5	Die drei unterschiedlichen Distanzen . . . . .	33
4.6	Die vier unterschiedlichen Längen . . . . .	35
4.7	Verschiedene Dichten . . . . .	36
4.8	Parallele Koordinaten mit und ohne Gap-Filling . . . . .	36
4.9	Benutzeroberfläche des Programmtools . . . . .	37
4.10	Sehtest . . . . .	37
4.11	Trainingtool: Erster Klick Falsch . . . . .	38

5.1	Genauigkeit in Abhängigkeit der Distanz: Die horizontale Achse stellt die Distanz und die vertikale Achse die Genauigkeit in Prozent dar. . . . .	39
5.2	Zeit in Abhängigkeit der Distanz: Die horizontale Achse stellt die Distanz und die vertikale Achse die Zeit in Sekunden dar . . . . .	40
5.3	Genauigkeit in Abhängigkeit der Linklänge: Die horizontale Achse stellt die Linklänge und die vertikale Achse die Genauigkeit in Prozent dar . . . . .	41
5.4	Zeit in Abhängigkeit der Linklänge: Die horizontale Achse stellt die Linklänge und die vertikale Achse die Zeit in Sekunden dar . . . . .	42
5.5	Genauigkeit in Abhängigkeit des Gap-Filling: Die horizontale Achse stellt das Gap-Filling und die vertikale Achse die Genauigkeit in Prozent dar . . . . .	43
5.6	Zeit in Abhängigkeit des Gap-Filling: Die horizontale Achse stellt das Gap-Filling und die vertikale Achse die Zeitt in Sekunden dar . . . . .	44
5.7	Genauigkeit in Abhängigkeit der Kombinationen Distanz und Gap-Filling: Die horizontale Achse stellt die Kombination von Distanz und GapFilling dar. Die vertikale Achse stellt die Genauigkeit in Prozent dar. . . . .	44
5.8	Genauigkeit in Abhängigkeit der Kombinationen Linklänge und Gap-Filling: Die horizontale Achse stellt die Kombination von Linklänge und GapFilling dar. Die vertikale Achse stellt die Genauigkeit in Prozent dar. . . . .	45
5.9	Genauigkeit in Abhängigkeit der Kombinationen Distanz und Linklänge: Die horizontale Achse stellt die Kombination von Distanz und Linklänge dar. Die vertikale Achse stellt die Genauigkeit in Prozent dar. . . . .	46
5.10	Zeit in Abhängigkeit der Kombinationen Distanz und Gap-Filling: Die horizontale Achse stellt die Kombination von Distanz und GapFilling dar. Die vertikale Achse stellt die Zeit in Sekunden dar. . . . .	47
5.11	Zeit in Abhängigkeit der Kombinationen Linklänge und Gap-Filling: Die horizontale Achse stellt die Kombination von Linklänge und GapFilling dar. Die vertikale Achse stellt die Zeit in Sekunden dar. . . . .	48
5.12	Zeit in Abhängigkeit der Kombinationen Distanz und Linklänge: Die horizontale Achse stellt die Kombination von Distanz und Linklänge dar. Die vertikale Achse stellt die Zeit in Sekunden dar. . . . .	49

# 1 Einleitung

In den letzten Jahrzehnten wurden bedeutende Forschungsanstrengungen zur Erforschung der Anwendbarkeit der Technik gewidmet und erweitert, was zu einer Vielzahl von Erweiterungen führt. Parallele Koordinaten sind seit vielen Jahren eine populäre Visualisierungstechnik für multivariate Daten. In den vergangenen Jahren stand die Technik im Fokus vieler Forschungsprojekte, vor allem in der Informationsvisualisierungsgemeinschaft, obwohl sie auch in der wissenschaftlichen Visualisierung und visuellen Analytik über zahlreiche Anwendungsgebiete wie Mathematik, Statistik, Bioinformatik und Medizin angewendet wurde.

Bereits im Jahre 1883 gab es Darstellungen von Parallelen Koordinaten durch H.Gannett und F.W.Hewes. Etwas später im Jahre 1959 wurde die ursprüngliche Idee von Alfred Inselberg erneut verwendet und von ihm im Jahre 1977 weiterentwickelt und popularisiert. Angewendet wurden sie am häufigsten bei Algorithmen zur Vermeidung von Zusammenstößen im Flugverkehr (1987), beim Data-Mining, bei Bildanalyseverfahren, in der Optimierung, der Prozesskontrolle, sowie der Einbruchserkennung bei Computern.

Eine Zusammenstellung der aktuellen Forschung zu Parallelkoordinaten findet sich im Stand der Technik von J.Heinrich und D.Weiskopf [2].

Die Visualisierung multivariater Daten kann bei Parallelen Koordinaten zu einer visuellen Unordnung führen, das heißt es entsteht eine Überlagerung und Kreuzung von vielen Kanten, welche die Analyse erschweren können. Die Erforschung von Parallelen Koordinaten konzentriert sich in erster Linie darauf, die visuelle Unordnung durch die Anzahl der Polylinien oder durch die Neuordnung der parallelen Achsen, zu verringern.

Gibt es ein Verfahren zur verbesserten Darstellung, d.h. um eine solche visuelle Unordnung zu reduzieren? In dieser Arbeit betrachten wir partielle Kanten in Parallelen Koordinaten um diese Unordnung zu reduzieren und stellen uns die Frage, ob solche Diagramme noch lesbar, verständlich und interpretierbar sind.

Das Ziel dieser Arbeit ist es, eine Benutzerstudie durchzuführen, um herauszufinden, ob partielle Kanten in Parallelen Koordinaten Vorteile gegenüber den normalen Parallelen Koordinaten haben.

Falls die Resultate der Studie zeigen sollten, dass sich gängige Aufgaben mit partiellen Kanten ebenso gut lösen lassen wie durchgehende Kanten, könnte dies Implikationen auf das Design von parallelen Koordinaten haben. Ein offensichtliches Beispiel wäre hier, dass der Raum zwischen Achsen genutzt werden könnte um zusätzliche Informationen anzuzeigen, welche die Betrachter bei der Lösung einer Aufgabe unterstützen könnten.

## Gliederung

Die Arbeit ist in folgender Weise gegliedert:

**Kapitel 2 – Grundlagen:** Hier werden werden die Grundlagen dieser Arbeit beschrieben.

**Kapitel 3 – Verwandte Arbeiten:** In diesem Teil werden verwandte Themen vorgestellt.

**Kapitel 4 – Studiendurchführung:** Der Ablauf der Studie wird beschrieben.

**Kapitel 5 – Auswertung:** In diesem Abschnitt werden die Ergebnisse der Studie genannt.

**Kapitel 6 – Zusammenfassung und Ausblick** fasst die Ergebnisse der Arbeit zusammen und stellt Anknüpfungspunkte vor.

# 2 Grundlagen

Dieses Kapitel beschäftigt sich mit den Grundlagen zum Thema der Bachelorarbeit. Das Kapitel erläutert verschiedene Visualisierungsarten sowie grundlegende Begriffe und Verfahren.

## 2.1 Multivariate Datenanalyse

Multivariate Datenanalyse beschäftigt sich mit der Suche nach Mustern und Beziehungen in Daten, die Variablen enthalten. Lineare Assoziation ist die einfachste Beziehung zwischen zwei Variablen. Korrelation dagegen, die am weitesten verbreitete Maßnahme zur Quantifizierung der Stärke und Richtung solcher bivariaten linearen Assoziationen. Die Korrelationsschätzung ist oft ein wichtiger erster Schritt in den meisten fortgeschrittenen Datenanalysen, welches wir auch in dieser Arbeit betrachten werden.

## 2.2 Visualisierungsarten für multivariate Daten

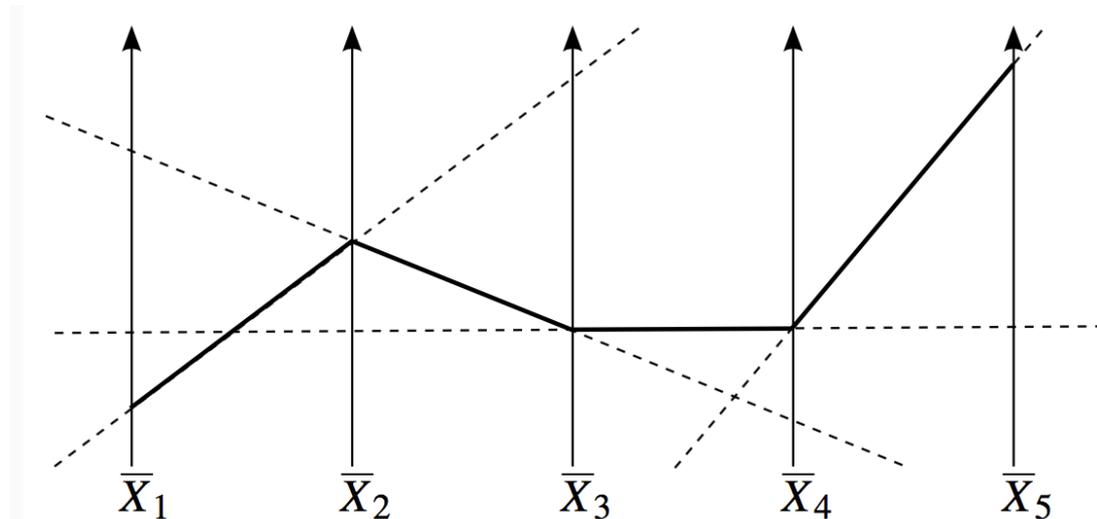
Im Laufe der letzten dreißig Jahre wurden zahlreiche Techniken zur Visualisierung von Informationen und Daten entwickelt und die Zahl der Visualisierungstechniken wird auch in Zukunft weiter ansteigen. Die beste Wahl eines Diagrammtyps hängt vom Zweck der Visualisierung ab. Da die Beschreibung aller Visualisierungstechniken den Rahmen dieser Arbeit sprengen würde, werden im Folgenden nur solche Ansätze erwähnt, die der Darstellung von multivariaten und multidimensionalen Datensätzen dienen.

### 2.2.1 Punkt- und Linienbasierte Darstellungen

#### Parallele Koordinaten

Eine Methode zur Visualisierung von hochdimensionalen Strukturen und multivariater Daten sind Parallele Koordinaten. Die Koordinatenachsen verlaufen parallel und in gleichem Abstand, siehe Abbildung 2.1.

Die Linien von links nach rechts entsprechen jeweils einem Datenpunkt und wird durch einen Polygonzug mit Ecken auf den parallelen Achsen dargestellt. Für jede gemessene Variable wird



**Abbildung 2.1:** Aufbau von parallelen Koordinaten mit fünf Dimensionen

eine vertikale Achse angelegt. Die Achse ist so skaliert, dass sie den gesamten Wertebereich der Daten abdeckt. Jeder Achsen-Schnittpunkt eines Linienzuges entspricht dem Wert des Datums bezüglich dieser Variable. Geschieht dies für alle Datentupel, erhält man die klassische parallele Koordinaten-Darstellung, siehe Abbildung 2.1. In der Arbeit von Heinrich und Weiskopf [2] werden Parallele Koordinaten ausführlich erklärt.

### Nachteile

Neben den vielen Vorteilen der Parallelen Koordinaten, gibt es natürlich auch Nachteile, die im folgenden beschrieben werden.

**Verdeckung:** Bei sehr vielen Datentupeln kommt es meist zur Verdeckung. Die gesamte Bildfläche erscheint einfarbig und es kann kaum Struktur abgelesen werden.

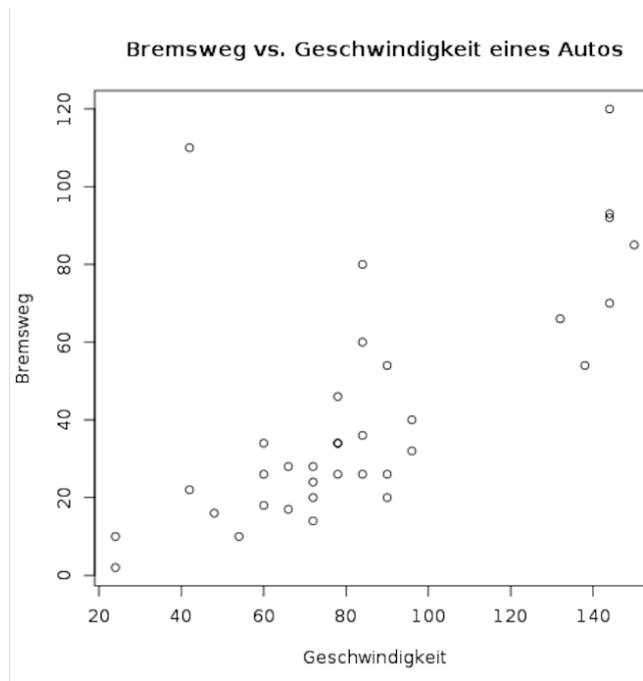
**Schwere Lesbarkeit:** Die Darstellung vieler Datentupel führt zu einer visuellen Unordnung.

**Unvollständige Darstellung:** Die Darstellung zeigt nur eine kleine Auswahl der paarweisen Korrelationen.

**Bias:** Die Anordnung der Achsen spielt eine sehr wichtige Rolle. Somit können Merkmale leichter erkannt werden.

## Streudiagramme

Streudiagramme zählen zu den bekanntesten Data Mining-Visualisierungstechniken und werden standardmäßig von vielen Statistik- und Tabellenkalkulationsprogrammen unterstützt. Streudiagramme stellen quantitative, bivariate Daten als Punktemengen dar. Sie sind für die Darstellung von zwei Variablen gut geeignet. Zwei Koordinatenachsen sind rechtwinklig zueinander angeordnet, wie bereits in Abbildung 2.2 zu sehen ist.



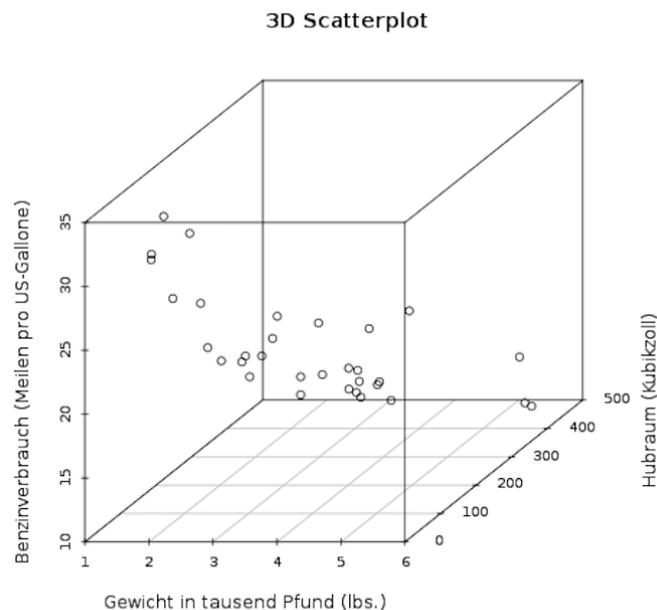
**Abbildung 2.2:** Bremsweg in Abhängigkeit der Ausgangsgeschwindigkeit: Die horizontale Achse stellt die Geschwindigkeit und die vertikale Achse den Bremsweg in einem Streudiagramm dar.

Hier wurde bei einigen Autos Vollbremsungen durchgeführt und die Ausgangsgeschwindigkeit und der resultierende Bremsweg gemessen. Auf dieser Abbildung fallen direkt drei Dinge auf:

- (1) Auto mit einer Geschwindigkeit zwischen 100 km/h und 130 km/h sind auf der Abbildung nicht zu sehen. Das kann an der Studie liegen, oder es ist ein Fehler bei der Datenverarbeitung unterlaufen.
- (2) Ein Auto hatte eine Ausgangsgeschwindigkeit von ca. 40 km/h, aber einen Bremsweg von 110 Metern, welches nur ein Ausreißer sein kann. Hier muss noch nachgeforscht werden. Gründe könnten sein, dass der Fahrer geschlafen oder der Versuchsleiter die Messungen in den Computer falsch eingegeben hat.

- (3) Der Zusammenhang wird hier besser durch eine Parabel als durch eine Gerade modelliert. Das macht Sinn, denn wenn man sich an die Fahrschule erinnert, ist der Bremsweg quadratisch abhängig von der Geschwindigkeit [1].

Streudiagramme für drei Variablen sind möglich, jedoch meist unübersichtlich. In der folgenden Abbildung wird als nächstes ein Datensatz von verschiedenen Autos visualisiert, in dem drei Variablen (Gewicht, Hubraum und Benzinverbrauch) dargestellt werden, siehe 2.3.



**Abbildung 2.3:** Streudiagramm für die drei Variablen Gewicht, Hubraum und Benzinverbrauch

Eine Darstellung bei mehr als drei Dimensionen ist nicht möglich. Auf eine grafische Darstellung von je zwei Variablen kann man jedoch zurückgreifen. Für jede Kombination von zwei Variablen werden somit Streudiagramme gezeichnet. Die bivariaten Streudiagramme werden dazu in einem rechteckigen Schema angeordnet, so dass jedes Merkmal einmal in der Zeile und einmal in der Spalte vorkommt. Die Merkmale der Zeilen werden auf der Abszisse (x-Achse) und die der Spalten auf der Ordinate (y-Achse) abgetragen. Sie unterscheiden sich in der Zuordnung der beiden Achsen.

Diese Darstellung wird Streudiagramm-Matrix genannt, siehe 2.4

Zusammenhänge können jedoch nicht mehr erkannt werden, die aus der Kombination von mehr als zwei Merkmalen resultieren.

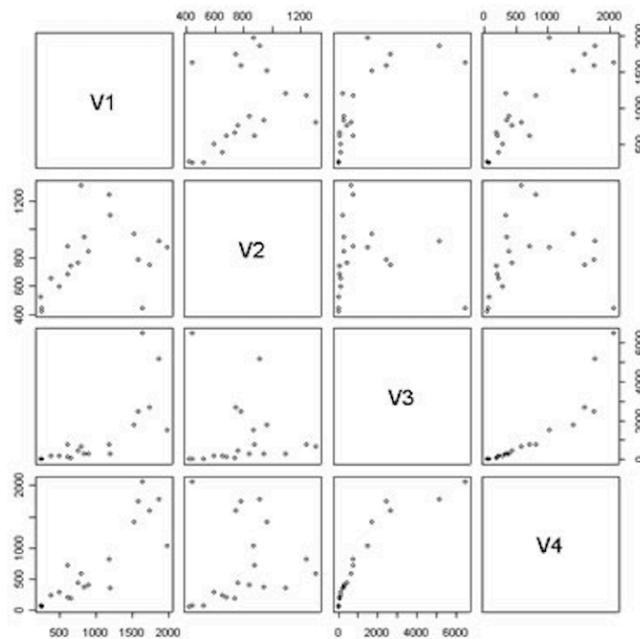


Abbildung 2.4: Beispiel: Mehrdimensional

### 2.2.2 Glyphen

Eine andere Klasse von explorativen Visualisierungstechniken stellen die sogenannten Icon- bzw. Glyph-basierten Ansätze dar. Diese Techniken werden, wie bei Parallelen Koordinaten und Streudiagrammen, zur Darstellung von diskreten, multivariaten Daten eingesetzt. Unter einem Glyphen versteht man dabei ein graphisches Objekt, welches ein einzelnes multivariates Datenobjekt darstellt. Bei der Generierung der Glyphen werden die diversen Datendimensionen eines Datensatzes den verschiedenen graphischen Attributen wie Form, Farbe, Größe, Orientierung, Textur, etc. des graphischen Objekts zugeordnet.

### Chernoff Gesichter

Chernoff-Gesichter zählen zu den am kunstvoll ausgearbeitesten Icon-Visualisierungsmethoden. Bei einer Visualisierung von Datensätzen mittels Chernoff-Gesichter wird jedem Datensatz eine comichafte Version eines menschlichen Gesichtes zugeordnet. Die einzelnen Elemente der Gesichter (Form, Augen, Mund, Nase, etc) werden jeweils einer Dimension zugeordnet, und dann je nach den Werten des Datensatzes verändert, siehe Abbildung 2.5.

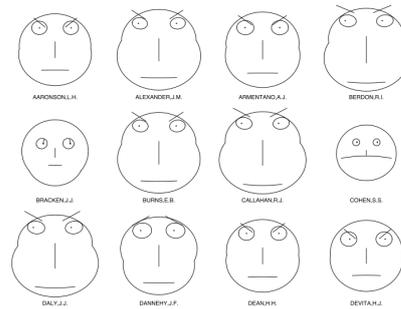


Abbildung 2.5: Zwölf verschiedene Chernoff-Gesichter

### Star Plots

In einem Star-Plot wird jeder Datensatz als eine sternähnliche Figur repräsentiert. Da die vorhandenen Dimensionen kreisförmig angeordnet werden, und die Werte des Datensatzes entlang jedes Strahls eingetragen werden, kommt die Sternform zustande, siehe Abbildung 2.6.

Jedes Datentupel eines Datensatzes kann entweder als eigenständiger Star-Plot dargestellt oder alle Tupel können gemeinsam in einem Plot zum Vergleich der Daten eingetragen werden.

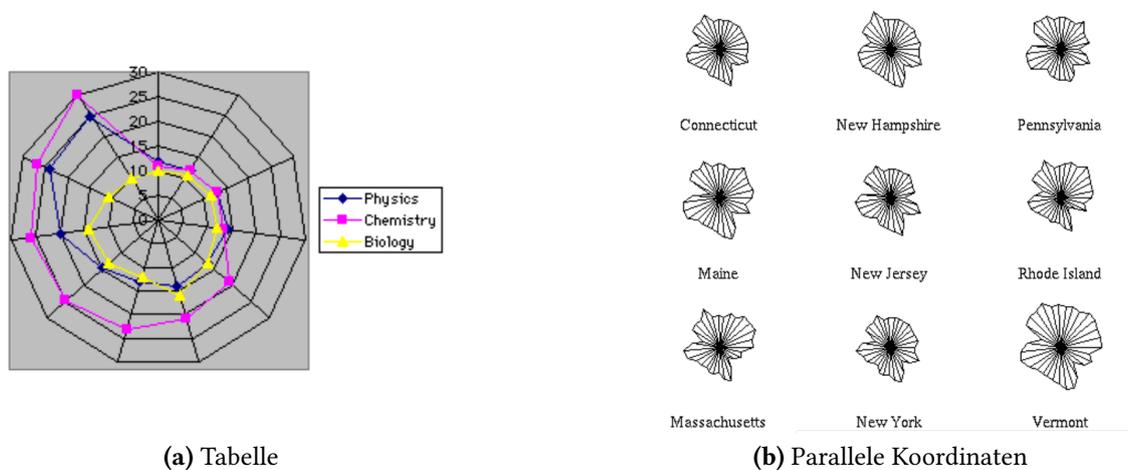


Abbildung 2.6: Beispiel: Star-Plots

## 2.3 Partielle Kanten

Partielle Kanten sind Linien, die teilweise gezeichnet werden anstatt jede Linie explizit in ihrer vollen Länge zu zeigen. Bei visuell kodierten Kanten zeichnen wir partielle Verbindungen, die

Vorne beginnen und auf das Ziel zeigen, anstatt dort genau zu enden. Dadurch kann die visuelle Unordnung reduziert werden, jedoch wird ein gewisses Maß an Unsicherheit eingeführt, welche sich auf die visuell nachvollziehbare Konnektivität von beispielsweise Knoten in einem Graphen bezieht.



**Abbildung 2.7:** Beispiel: Partielle Kanten in Parallelen Koordinaten

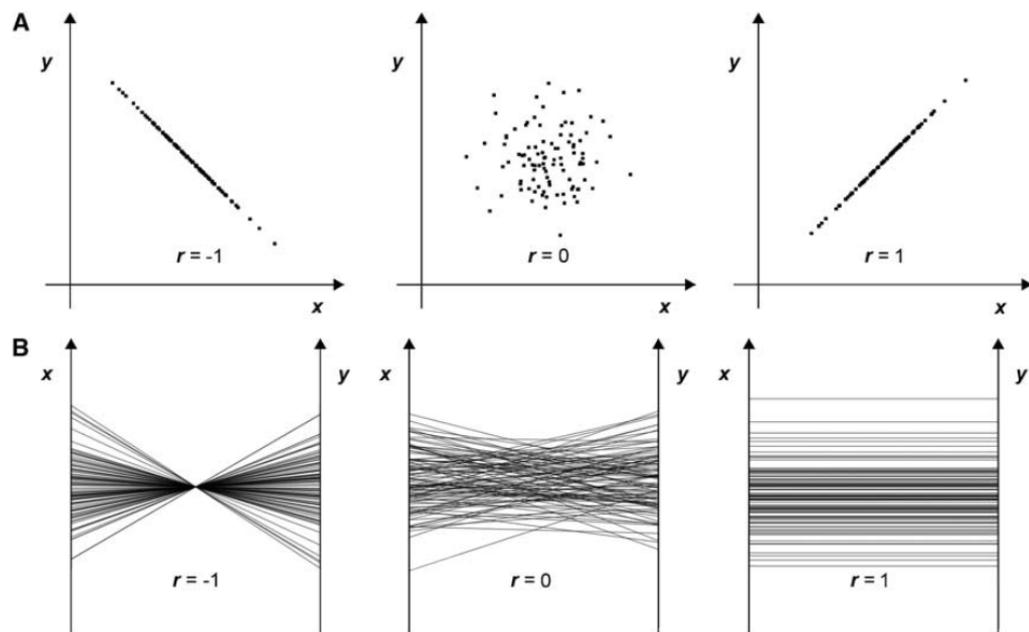
## 2.4 Korrelation

Eine Korrelation ist eine Beziehung zwischen mehreren Variablen. In der statistischen Analyse kann Korrelation auf verschiedene Weise definiert werden. Meist wird die Pearson Korrelation verwendet um die Korrelation zu quantisieren.

Der Korrelationskoeffizient für zwei quadratisch integrierbare Zufallsvariablen  $X$  und  $Y$  mit jeweils positiver Standardabweichung  $\sigma_X$  und  $\sigma_Y$  und Kovarianz lautet:

$$\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y}$$

Der Korrelationskoeffizient ist ein Maß für den Grad des linearen Zusammenhangs zwischen zwei intervallskalierten Merkmalen. Die Pearson Korrelation kann Werte zwischen -1 und +1 annehmen. Bei einem Wert von +1 (bzw. -1) besteht ein positiver (bzw. negativer) linearer Zusammenhang zwischen den betrachteten Merkmalen. Wenn der Korrelationskoeffizient 0 aufweist, hängen die beiden Merkmale überhaupt nicht linear voneinander ab. Positive



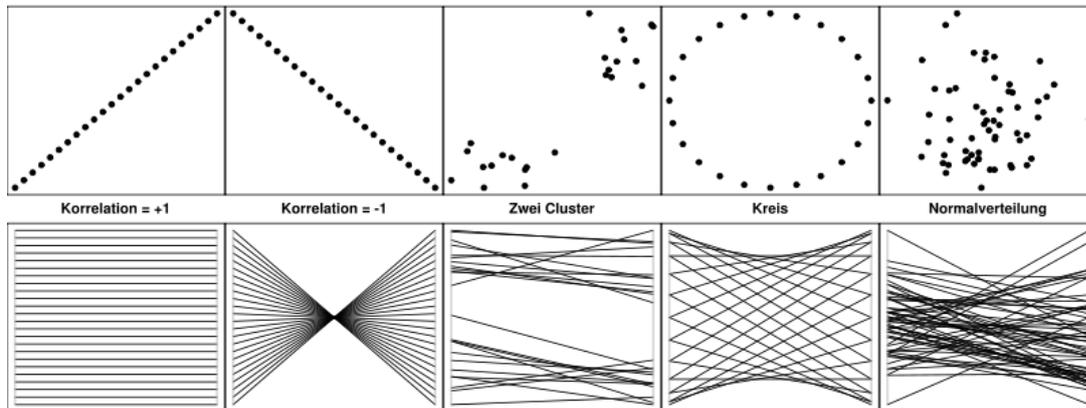
**Abbildung 2.8:** Drei Extremwerte der Korrelation für die Visualisierungsarten (A) Streudiagrammen und (B) Parallele Koordinaten

Korrelation führt zu parallelen Linien. Bei der Negativen Korrelation schneiden sich die Linien in einem Punkt, siehe Abbildung 2.8.

Einige Beispiele für Streudiagramme und korrespondierenden Mustern in parallelen Koordinaten sind in Abbildung 2.9 dargestellt.

Aus dem Diagramm kann man folgende Informationen direkt ablesen:

- Wie sind die Daten innerhalb einer Variablen verteilt?
- Gibt es Werte die häufig/selten auftreten?
- Welche Variablen sind kontinuierlich, welche diskret?
- Welche Zusammenhänge bestehen mit "benachbarten" Variablen?
- Viele gleich gerichtete Linien weisen auf eine positive Korrelation, viele sich kreuzende Linien auf eine negative Korrelation hin.
- Wie groß ist die Streuung bei Verbindungen zwischen "benachbarten" Variablen?



**Abbildung 2.9:** Vergleich Streudiagramm und Parallele Koordinaten



## 3 Verwandte Arbeiten

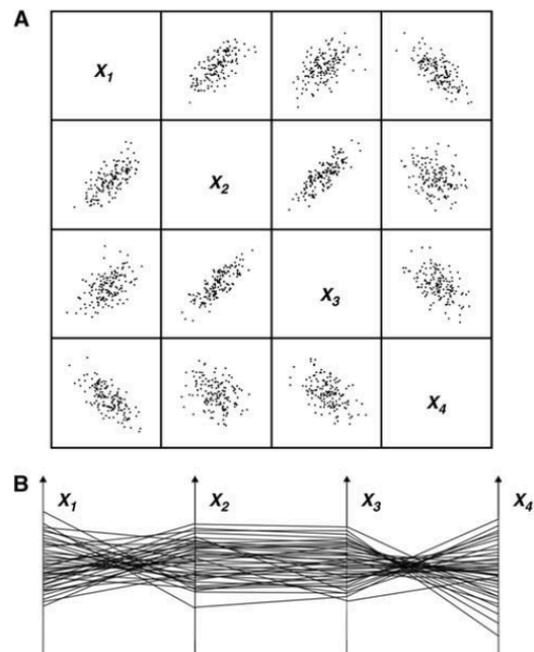
In diesem Kapitel wird auf verwandte Themen eingegangen, die zum Verständnis dieser Bachelorarbeit nötig sind und die den aktuellen Forschungsstand aufzeigen.

### 3.1 Vergleich Streudiagramme und Parallele Koordinaten

In der Forschungsarbeit “Judging correlation from scatterplots and parallel coordinate plots” von Jung Li, Jean-Bernard Martens und Jarke J van Wijk steht der Vergleich der beiden Visualisierungsmethoden, Streudiagramme und Parallele Koordinaten, im Blickpunkt. Streudiagramme und Parallele Koordinaten können verwendet werden um die Korrelation visuell zu beurteilen, siehe Abbildung 3.1 In einer Studie wurden 25 Teilnehmer aufgefordert, die Korrelation unter verschiedenen Bedingungen der Visualisierungsmethoden, Datengröße und Beobachtungszeit zu schätzen. Das Ziel der Arbeit ist es, die Wirksamkeit dieser beiden Visualisierungsmethoden für die Korrelationsanalyse zu bewerten und daraus herauszufinden, wie die Versuchspersonen mit beiden Visualisierungsmethoden umgehen. Die visuellen Stimuli im Experiment werden durch die Datensätze, das Visualisierungsverfahren  $V$  (kursiv) und die Anzeigzeitbedingung  $T$  (kursiv) bestimmt. Die Studie begann mit einem Tutorial, das Scatterplots und PCPs erklärt hat und wie man sie benutzt, um die Korrelation zu analysieren. Bilder beider Visualisierungsmethoden für  $r = -1$ ,  $r = 0$  und  $r = 1$  wurden auf einem Papier dargestellt. Als nächstes gab es eine Testversion, in der sich die Teilnehmer mit der Testumgebung und der Testschnittstelle vertraut machen konnten. Danach begann die richtige Studie. Für jede Sitzung wurden die 42 Stimuli, die durch die Kombination von vorgenerierten Datensätzen und Visualisierungsbedingungen erzeugt wurden, in zufälliger Reihenfolge dargestellt. Schließlich konnten die Teilnehmer ein Feedback geben. Die Graphen wurden auf dem PC-Bildschirms angezeigt, siehe Abbildung 4.9. Basierend auf Analysen konnten sie schlussfolgern, dass bei der visuellen Korrelationsanalyse Streudiagramme viel effektiver als Parallele Koordinaten sind.

### 3.2 Partielle Kanten in gerichteten Graphen

In der früheren Arbeit [becker] beschäftigte sich Becker bereits mit partiellen Kanten. Becker hat keine Studie für die Auswirkungen partieller Kanten durchgeführt.



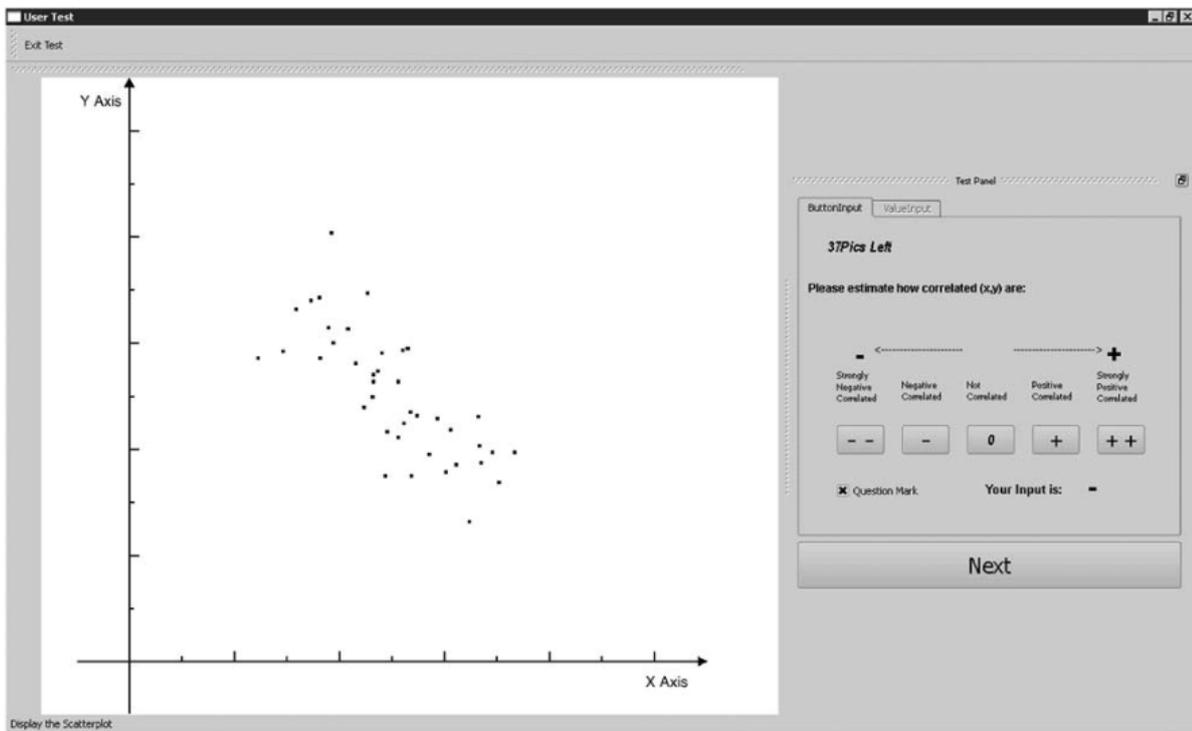
**Abbildung 3.1:** Darstellung multivariater Daten: (a) Streudiagramm Matrix und (b) Parallele Koordinaten

### 3.3 Partielle Kanten in gerichteten Graphen

Partielle Kanten wurden auch in gerichteten Graphen angewendet um die visuelle Unordnung zu reduzieren [burch]. Eine Studie mit 42 Teilnehmern wurde durchgeführt um die Frage, ob solche Graphen noch lesbar, verständlich und interpretierbar sind, zu beantworten. In der Studie wurden konische und traditionelle Kantendarstellungen, drei verschiedene Graphengrößen und sechs verschiedene Längen untersucht. Ein Ergebnis dieser Studie ist, dass die Merkmale der Bearbeitungszeiten und die Fehlerraten von der Aufgabenart abhängig sind.

### 3.4 Übersicht Paralleler Koordinaten

In dieser Forschungsarbeit [burch] beschäftigten sich Jimmy Johansson und Camilla Forsell mit Parallele Koordinaten und stellen eine Übersicht, Kategorien und Richtlinien für die



**Abbildung 3.2:** Benutzeroberfläche des Programmtools

zukünftige Forschung zusammen. Basierend auf die Daten einer Umfrage und der Erfahrungen der Autoren, die mit parallelen Koordinaten arbeiten, wird ein Leitfaden für zukünftige Forschungsrichtungen vorgeschlagen.



# 4 Studiendurchführung

Die Studie befasst sich hauptsächlich mit der Schätzung der Korrelation bei parallelen Koordinaten mit partiellen Kanten. Im Vergleich zu normalen Kanten ist nun in der Mitte der Schnitt der Kanten nicht ersichtlich. In der Studie werden zweidimensionale Parallele Koordinaten mit verschiedenen Parametern betrachtet, um die Schwierigkeit der Schätzung genau auszuwerten.

In diesem Kapitel wird das Studiendesign vorgestellt. Anschließend wird auf die Stimuli, die Aufgabenbearbeitung, die Pilotstudie, die Untersuchungsumgebung und die Versuchspersonen eingegangen, bevor der eigentliche Studienablauf ins Blickfeld tritt.

## 4.1 Studiendesign und Parameter

Unter dem Studiendesign wird die Gesamtheit der Vorgehensweisen im Rahmen einer Studie verstanden. Ein gutes Studiendesign versucht Einflüsse, die das Ergebnis eines Testverfahrens verfälschen können, auszuschließen.

Das Experiment beinhaltet nur eine Aufgabe, nämlich das Schätzen der Korrelation bei Parallelen Koordinaten. Für die Studie betrachten wir folgende Parameter.

- (1) Länge der Kanten
- (2) Distanz der Achsen
- (3) Mit oder Ohne Gap-Filling
- (4) Datenanzahl

Zuerst wird eine Pilotenstudie mit zehn Versuchspersonen durchgeführt um eventuell bestehende Mängel in der Aufgabenstellung oder bei der Durchführung der Studie aufzudecken und anschließend zu beseitigen. Danach wird die eigentliche Studie mit 36 Teilnehmern durchgeführt, welche mit einer Trainingsphase beginnt.

### 4.2 Pilotenstudie

Zu Beginn wird eine Pilotenstudie durchgeführt, um zu entscheiden, welche Parameter für die richtige Studie in Betracht gezogen werden können und welche nicht relevant sind. Auch um eventuell bestehende Mängel in der Aufgabenstellung oder bei der Durchführung der Studie aufzudecken und anschließend zu beseitigen. An der Pilotstudie nahmen zehn Versuchspersonen teil, acht von ihnen waren weiblich und zwei männlich. Nur einer der Versuchspersonen hatte Vorkenntnis im Bereich der Visualisierung. Bei der Pilotstudie betrachteten wir folgende zwei Gruppen, wobei es kein Gap-Filling gab, weil dies sowieso bei der eigentlichen Studie verwendet wird:

- (1) Variable Linklänge: Die Distanz zwischen den Achsen ändert sich nicht. Für die Linklänge testen wir die Werte 0.1, 0.2, . . . , 0.8, wobei eine Zahl der Anteil der Distanz zwischen den Achsen ist.
- (2) Variable Distanz: Die Werte für die Linklänge bleiben gleich. Für die Distanz zwischen den Achsen testen wir die Werte 50, 100, 150, 200, wobei die Zahl die Anzahl der Pixel zwischen den Achsen ist.

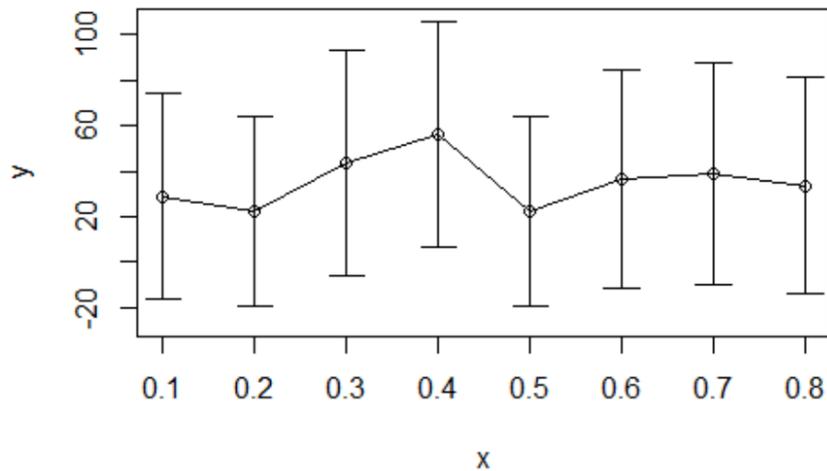
Die Hälfte der Versuchspersonen führt die Studie mit variabler Linklänge durch und die andere Hälfte mit variabler Distanz.

#### 4.2.1 Resultat Linklänge

Bei Länge-Genauigkeit sieht man, dass die Mittelwerte ab 0.2 langsam ansteigen. Die Streuung der Messdaten bleibt konstant. Ab 0.4 treten Schwankungen auf und ab 0.6 fallen die Mittelwerte ab, siehe Abbildung 4.1. Bei Länge-Zeit. Schwankungen von 0.1 bis 0.4 mit anschließendem Abfall der Mittelwerte bis 0.6. Streuung der Messdaten schwankt zu Beginn und verringert sich ab 0.4 bis 0.6, siehe Abbildung 4.2. Basierend auf den Beobachtungen zu Länge-Zeit-Genauigkeit wurden folgende Längeparameter als relevant befunden und in der Hauptstudie verwendet: 0.1, 0.3, 0.6, 0.8.

#### 4.2.2 Resultat Distanz

Bei Distanz-Genauigkeit sieht man, dass die Mittelwerte langsam ansteigen. Die Streuung der Messdaten quasi konstant ist. Bei 300 treten Schwankungen auf, siehe Abbildung 4.3. Bei Distanz-Zeit: Abfall von 50 bis 150 der Mittelwerte. Streuung der Messdaten verringert sich auch. Plateau von 150 bis 250. Anstieg bei 300 mit anschließendem Abfall der Mittelwerte. Streuung nimmt auch wieder zu, siehe Abbildung 4.4 Basierend auf den Beobachtungen zu Distanz-Zeit-Genauigkeit wurden folgende Distanzparameter als relevant befunden und in der Hauptstudie verwendet: 50, 150 und 300 pixel.



**Abbildung 4.1:** Genauigkeit in Abhängigkeit der Linklänge: Die horizontale x- Achse stellt die relative Linklänge und die vertikale y-Achse die Genauigkeit in Prozent dar.

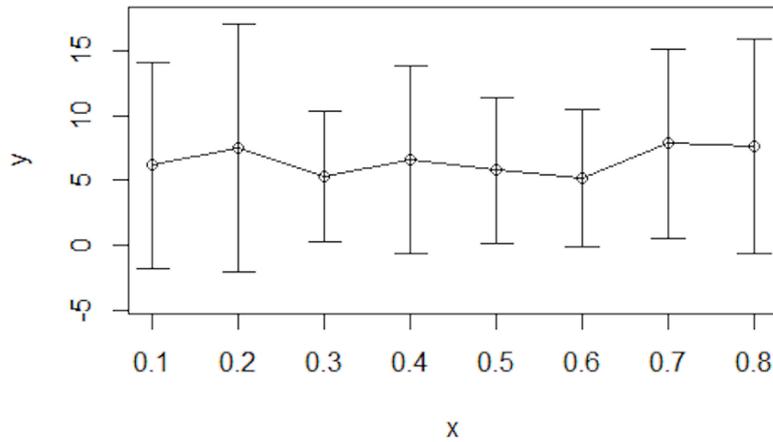
### 4.2.3 Resultate

Im Durchschnitt brauchte ein Teilnehmer circa sechs Sekunden, um die Korrelation bei einem Bild zu schätzen. In den Diagrammen ist zu sehen, wie das Verhältnis zwischen Länge-Zeit und Länge-Korrektheit ist. Auch zwischen Distanz-Zeit und Distanz-Korrektheit. Wir haben gesehen, dass wir nicht alle Parameter und Paare testen konnten, da die Teilnehmer für 41 Stimuli circa 6 Sekunden brauchten. Um einen Effekt zu erzielen, begrenzten wir die Parameterwerte. Somit wurden die Werte der Länge 0.1, 0.3, 0.4, 0.6 und 0.8 verwendet. Für die Werte der Distanz 50,150,300 und 400.

## 4.3 Hypothese

Wir spekulieren, dass partielle Kanten in Parallelen Koordinaten zu mehr graphischen Fehlinterpretationen führen kann, sind uns aber unsicher, wie sich Fehlerraten und Bearbeitungszeiten bei der Lösung von grafischen Aufgaben verhalten. Basierend auf den Beobachtungen der Pilotstudie wurden folgende Hypothesen aufgestellt und getestet.

- H1: Geringere Linklänge führt zu geringer Genauigkeit und längerer Bearbeitungszeit.
- H2: Große Distanz zwischen zwei Achsen führt zu einer längeren Bearbeitungszeit sowie höhere Fehlerraten.



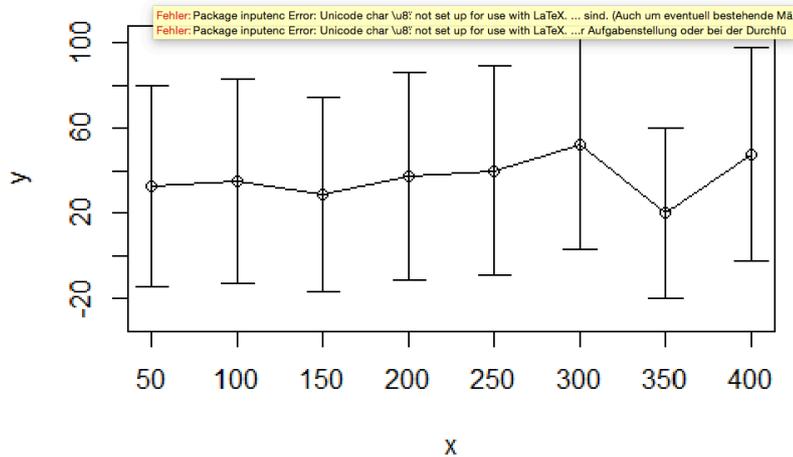
**Abbildung 4.2:** Zeit in Abhängigkeit der Linklänge: Die horizontale x- Achse stellt die relative Linklänge und die vertikale y-Achse die Zeit in Sekunden dar.

H3: Bei der Darstellung mit Gap-Filling spekulieren wir eine hohe Fehlerrate, da es irreführend ist.

H1: Geringere Linklänge führt zu geringer Genauigkeit und längerer Bearbeitungszeit. H2: Große Distanz zwischen zwei Achsen führt zu einer längeren Bearbeitungszeit sowie höhere Fehlerraten. H3: Bei der Darstellung mit Gapfilling spekulieren wir eine hohe Fehlerrate, da es irreführend ist.

### 4.4 Stimuli

360 Stimuli wurden durch die Kombination von vorgenerierten Datensätzen mit RStudio in zufälliger Reihenfolge angezeigt. Befragt wurden 36 Teilnehmer. Für jede Session wurden die 36 Stimuli, die durch die Kombination von vorgenerierten Datensätzen in RStudio erzeugt wurden, in zufälliger Reihenfolge dargestellt. Die Versuchspersonen erhalten 360 Bilder, bei welchen Sie die Korrelation schätzen müssen. Als Parameter wurde die Länge der Kanten, die Distanz der Achsen und die Darstellungsvariante mit und ohne Gapfilling verwendet. In den Abbildungen 4.5, 4.6, 4.7 und 4.8 sind die unterschiedlichen Stimuli zu sehen.



**Abbildung 4.3:** Distanz-Genauigkeit: Genauigkeit in Abhängigkeit der Distanz: Die horizontale x-Achse stellt die Distanz zwischen den Achsen in Pixeln dar und die vertikale y-Achse die Genauigkeit in Prozent.

## 4.5 Rahmenwerk

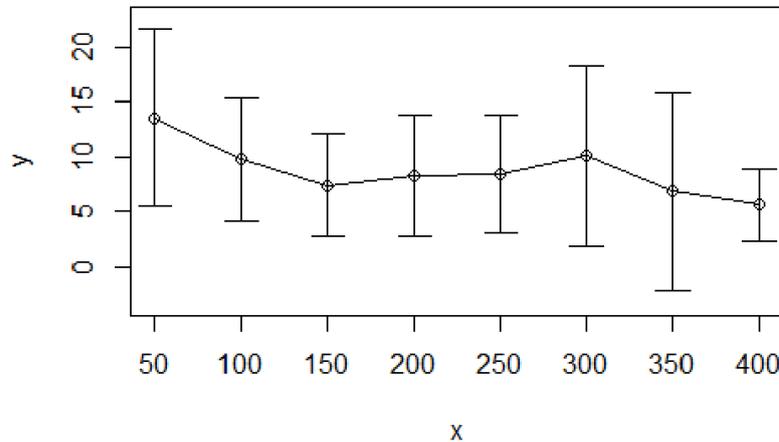
Das Programmtool kann über das Terminal geöffnet werden. Sobald sich das Fenster öffnet erscheint ein grauer Hintergrund mit dem Button Start in der Mitte. (Foto noch hinzufügen). Sobald die Versuchsperson bereit ist, kann sie auf den Start-Button klicken und mit der Studie beginnen. Es erscheinen nun die Bilder auf einem weißen Hintergrund. Unter dem Bild sind fünf Buttons für die fünf Korrelationen zu sehen, siehe Abbildung 4.9

## 4.6 Technische Daten

Die Graphen wurden auf der linken Seite des PC-Bildschirms in einem Bereich von 24 × 26 cm mit weißem Hintergrund und schwarzen Objekten angezeigt. Im Kapitel () wird das Rahmenwerk, dass zur Studiendurchführung implementiert wurde, genau beschrieben.

## 4.7 Teilnehmer

36 Teilnehmer haben an der Studie teilgenommen. Teilgenommen haben Studenten, Auszubildende unter anderem welche aus unterschiedlichen Branchen, etwa dem Handwerk und dem Kaufmännischen Bereich im Alter von 18-36 Jahren. Davon waren 20 Studenten mit



**Abbildung 4.4:** Distanz-Zeit: Dauer bis die Korrelation von den Teilnehmern bestimmt wurde in Abhängigkeit der Distanz. Die horizontale Achse ist die Distanz zwischen den Achsen in Pixeln und die vertikale Achse die Zeit in Sekunden

dem Abschluss Bachelor. Mit dem Begriff Parallele Koordination waren nur 5 der Teilnehmer vertraut. Alle arbeiten wöchentlich am Computer und haben eine normale Sehkraft, was über einen Snellen-Index bestätigt wurde. Die Studie dauerte im Schnitt 50 Minuten. Die schnellste Versuchsperson benötigte 40 Minuten und der langsamste 70 Minuten.

## 4.8 Studienablauf

Die Studie konnte sowohl an der Universität Stuttgart im Informatikgebäude am Rechner als auch an einem Laptop durchgeführt werden. In den folgenden Unterkapiteln wird der Ablauf der Studie genauer geschildert, beginnend mit dem Unterschreiben einer Einverständniserklärung, dem Ausfüllen eines Fragebogens zur Person, die Durchführung eines Sehtests, Das Durchführen eines Tutorials. Im Anschluss wurde die Hauptstudie durchgeführt.

### 4.8.1 Einverständniserklärung

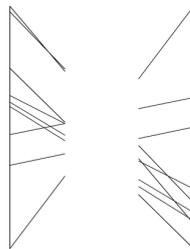
Zu Beginn der Studie bekommen die Versuchspersonen eine Einverständniserklärung. Diese enthält eine Erläuterung zum Ablauf der Studie sowie den Hinweis, dass die Studie jederzeit abgebrochen oder pausiert werden kann. Des Weiteren wird darauf hingewiesen, dass die gesammelten Daten anonym behandelt werden.



(a) Distanz 50



(b) Distanz 150



(c) Distanz 300

**Abbildung 4.5:** Die drei unterschiedlichen Distanzen

### 4.8.2 Fragebogen zum Probanden

Der Fragebogen beinhaltet Fragen zu Geschlecht, Alter, Muttersprache, Sehhilfe, Computerarbeitszeiten und Abschluss.

### 4.8.3 Sehtest

Um sicherzustellen, dass die Probanden keine Sehschwäche aufweisen, die die Ergebnisse der Studie beeinträchtigen könnten, wird ein Sehtest durchgeführt. Der Test wird mithilfe eines SSnellen Chart" durchgeführt, das auch Augenärzte einsetzen um die Sehschärfe zu bestimmen. Wir nutzen diesen Test zwar, geben den Teilnehmern aber keinen medizinischen Rat oder Diagnose. Alle Teilnehmer hatten eine normale Sehschärfe gegeben falls unter Verwendung einer Sehhilfe. Der Test hängt an der Wand und der Abstand beträgt 3 Meter. Der Teilnehmer muss die Zahlen beginnend mit der oberen Reihe vorlesen, siehe Abbildung 4.10

### 4.8.4 Tutorial

Bevor die einzelnen Aufgaben bearbeitet werden, erhalten die Versuchspersonen ein Blatt, in dem zu Beginn eine allgemeine Beschreibung enthalten ist. Darauf sind drei verschiedene Möglichkeiten der Korrelation in Parallelen Koordinaten dargestellt. Danach besteht die Möglichkeit, eventuelle Fragen zur Korrelation oder zum Ablauf der Studie zu stellen. Als nächstes gab es eine Test-Version, in der sich die Teilnehmer mit der Testumgebung und der Testschnittstelle vertraut machen konnten, siehe Abbildung 4.11

Wenn sie auf ein Button klicken und dieser ist falsch, dann färbt sich dieser Button rot. Hat der Teilnehmer jedoch auf den richtigen Button geklickt, dann geht es direkt weiter. Die Teilnehmer konnten ca. 10 Bilder testen bis sie sich in der Thematik sicher gefühlt haben. Es durften erneut Fragen gestellt werden. Nach dieser Trainingsphase, startet die eigentliche Studie.

### 4.8.5 Durchführung der Aufgaben

Die eigentliche Studie umfasst 360 Stimuli. Die Studie beinhaltet nur einen Aufgabenblock. Im Programm wird beschrieben, was die Versuchspersonen bei der Aufgabe zu erledigen haben, siehe Abschnitt 4.5. Hier färbt sich kein Button mehr, das heißt der Teilnehmer erfährt nicht, ob seine Ergebnisse richtig sind, sondern klickt sich durch die 360 Bilder. Zu Beginn wird dem Teilnehmer mitgeteilt so schnell und so gut wie möglich. Während der Studie dürfen keine Fragen mehr gestellt werden. Im Hintergrund läuft ein „Timer“, welchen der Teilnehmer nicht sehen kann. Das Fenster verschwindet, sobald der Teilnehmer die 360 Bilder ausgeführt hat. Anschließend musste der Teilnehmer ein Fragebogen ausfüllen. Für die Teilnahme erhielt jeder eine Tafel Schokolade.



(a) Länge 0.1



(b) Länge 0.3



(c) Länge 0.6



(d) Länge 0.8

**Abbildung 4.6:** Die vier unterschiedlichen Längen

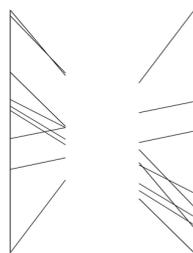


(a) wenige Daten

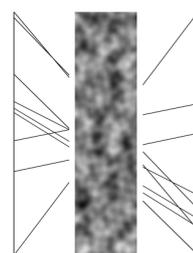


(b) viele Daten

**Abbildung 4.7:** Verschiedene Dichten

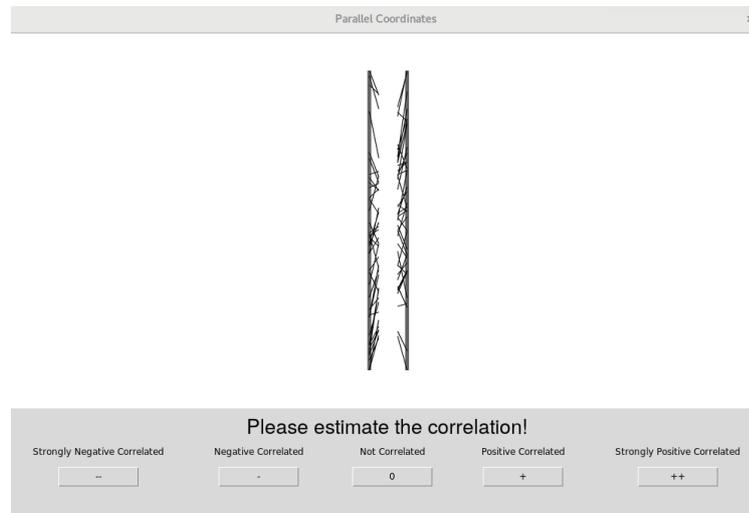


(a) Distanz 50



(b) Distanz 150

**Abbildung 4.8:** Parallele Koordinaten mit und ohne Gap-Filling



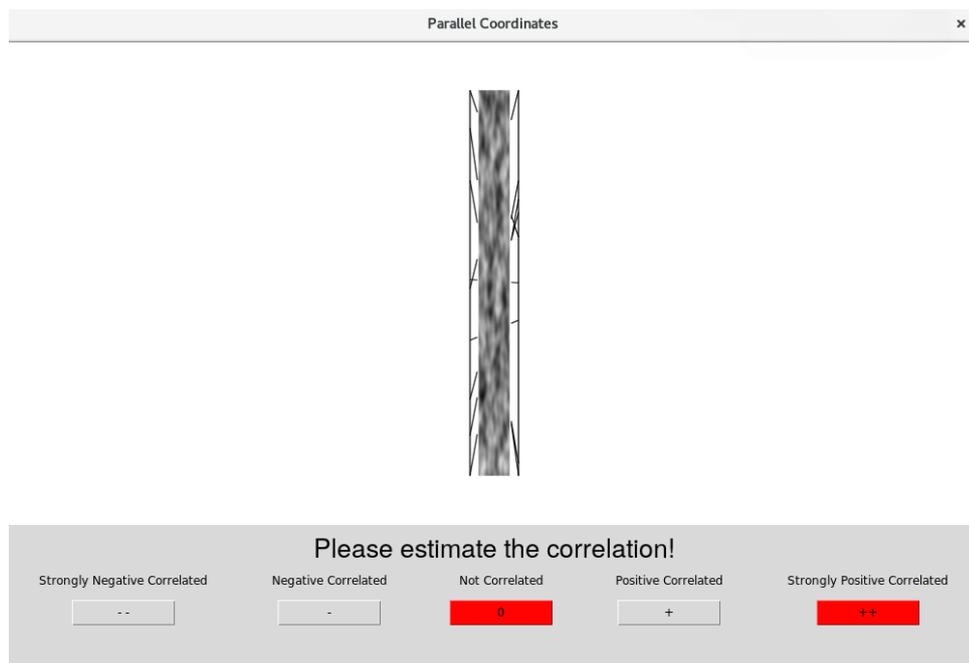
**Abbildung 4.9:** Benutzeroberfläche des Programmtools

Snellen test

Zeilen, die nicht gelesen werden konnten:

- 1 20/200
- 2 20/100
- 3 20/70
- 4 20/50
- 5 20/40
- 6 20/30
- 7 20/25
- 8 20/20
- 9
- 10
- 11

**Abbildung 4.10:** Sehtest

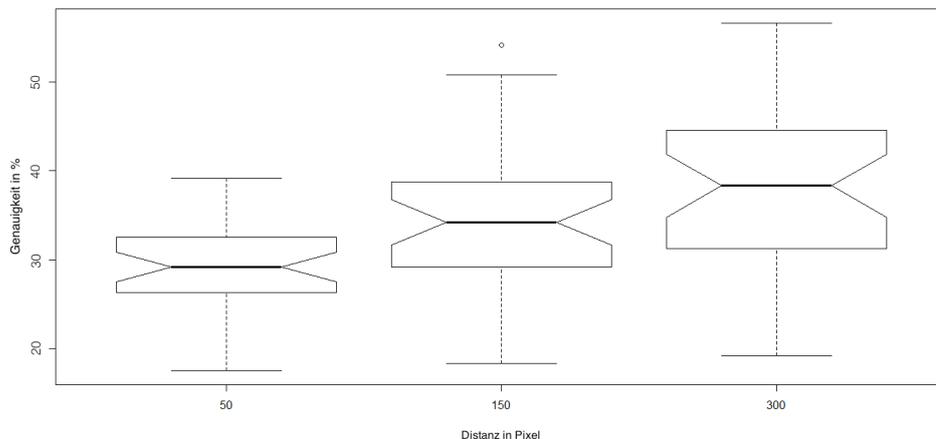


**Abbildung 4.11:** Trainingtool: Erster Klick Falsch

# 5 Auswertung

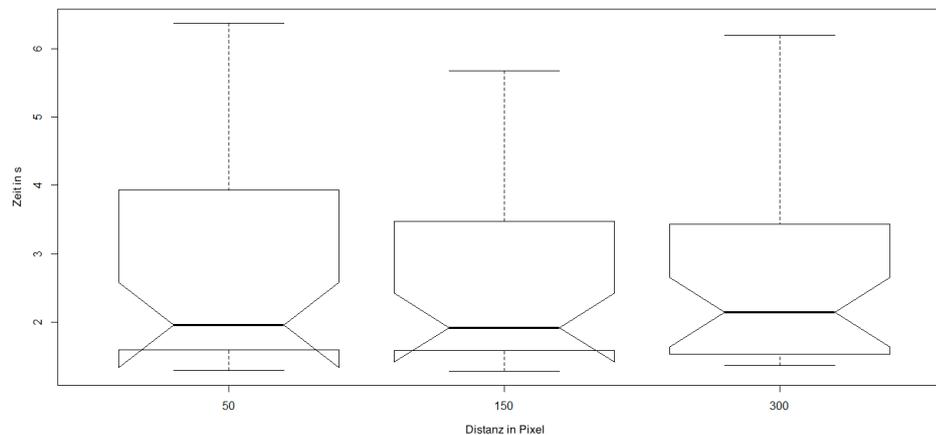
Dieses Kapitel stellt die Auswertungsergebnisse dar. Das erste Unterkapitel stellt die Ergebnisse bei variabler Distanz dar und das darauffolgende die Ergebnisse mit variabler Länge. In beiden Fällen wurde die Bearbeitungszeit und die Genauigkeit untersucht.

## 5.1 Distanz



**Abbildung 5.1:** Genauigkeit in Abhängigkeit der Distanz: Die horizontale Achse stellt die Distanz und die vertikale Achse die Genauigkeit in Prozent dar.

Die Distanz scheint eine Auswirkung auf die Genauigkeit zu haben unter Betrachtung des dazugehörigen Boxplots (siehe Bild 5.1). Größere Distanz bedeutet in der Regel bessere Korrektheit. Die Teilnehmer haben am besten bei einer Distanz von 300 abgeschnitten, mit knapp 40% gefolgt von einer Distanz von 150 mit 33% und schließlich mit Distanz 50 und Genauigkeit 29%. Laut der statistischen Auswertung liegen signifikante Differenzen vor:



**Abbildung 5.2:** Zeit in Abhängigkeit der Distanz: Die horizontale Achse stellt die Distanz und die vertikale Achse die Zeit in Sekunden dar

( $\chi^2 = 19.823$ ;  $p = 4.96e - 05$ ) Wir erklären es uns so: Mit steigender Distanz vergrößern sich die Linklängen, da sie relativ zum Abstand der Parallelen definiert werden. Die Post-Hoc-Analyse zeigt, dass folgendes Distanzpaar signifikante Differenzen aufweist: 50-150.

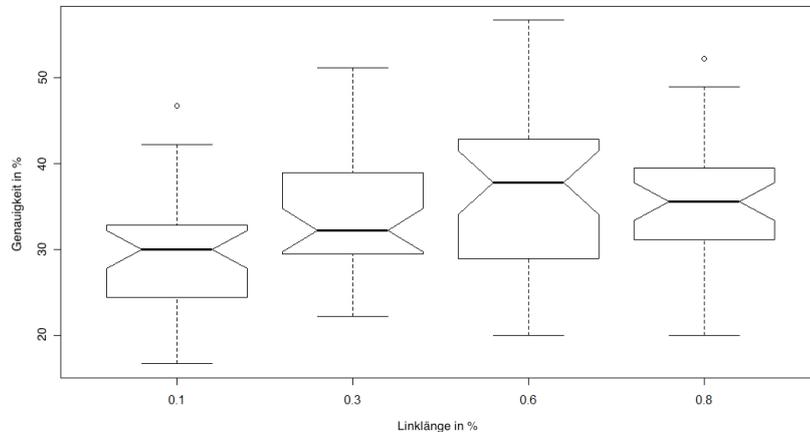
Zeit scheint auf den ersten Blick eine Rolle zu spielen unter Betrachtung des Boxplots (siehe Bild 5.2). Laut Statistiktest liegen keine signifikanten Differenzen vor: ( $\chi^2 = 0.021132$ ;  $p = 0.9895$ ).

Anhand der Testergebnisse kann Hypothese H2, dass eine große Distanz zwischen zwei Achsen zu einer längeren Bearbeitungszeit sowie einer höheren Fehlerrate führt, nicht bestätigt werden.

## 5.2 Länge

Unter Betrachtung des Boxplots (siehe Bild 5.11) scheint Zeit nicht relevant zu sein und Genauigkeit scheint signifikante Unterschiede zu haben. Größere Linklänge bedeutet in der Regel bessere Korrektheit. Die Teilnehmer haben am besten bei einer Linklänge von 0.6 abgeschnitten, mit knapp 40% gefolgt von einer Länge von 0.8 mit 35% und die Linklänge 0.3 mit 31% und schließlich mit der Länge 0.1 und Genauigkeit 30%.

Dies wird durch den Statistiktest bestätigt: ( $\chi^2 = 15.507$ ;  $p = 0.001431$ )



**Abbildung 5.3:** Genauigkeit in Abhängigkeit der Linklänge: Die horizontale Achse stellt die Linklänge und die vertikale Achse die Genauigkeit in Prozent dar

Die Post-Hoc-Analyse zeigt, dass folgendes Linklängen-Paar signifikante Differenzen aufweisen: 0.1-0.6; 0.1-0.8. Das heißt die optimale Linklänge liegt in der Umbegung von 0.6-0.8, da die Mittelwerte vor 0.6 stetig steigen aber bei 0.8 wieder fallen, daher könnte ein Maximum zwischen 0.6 und 0.8 liegen.

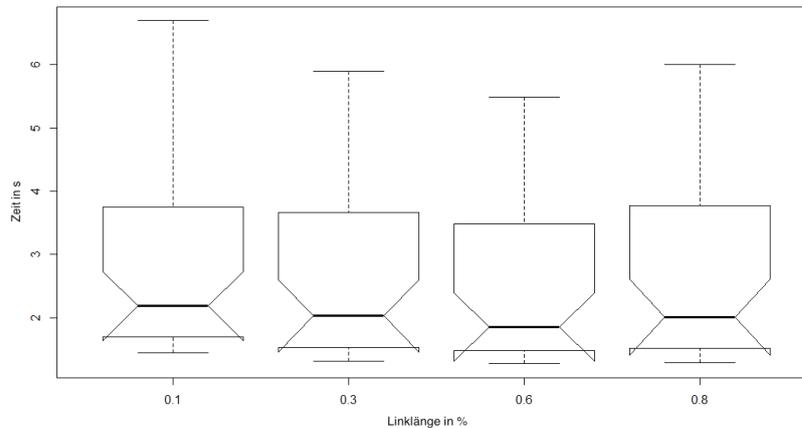
Die Hypothese H1, dass geringere Linklänge zu geringer Genauigkeit, kann bestätigt werden.

## 5.3 Gap

Für beide sieht man anhand der Boxplots 5.5 und 5.6 nicht viel.

Laut der statistischen Auswertung liegen für beide keine signifikante Differenzen vor: ( $\chi^2 = 0.58377$ ;  $p = 0.4448$ ) und ( $\chi^2 = 0.076194$ ;  $p = 0.7825$ )

Die passende Hypothese H3, dass wir bei der Darstellung mit Gapfilling, eine hohe Fehlerrate spekulieren, kann nicht bestätigt werden.



**Abbildung 5.4:** Zeit in Abhängigkeit der Linklänge: Die horizontale Achse stellt die Linklänge und die vertikale Achse die Zeit in Sekunden dar

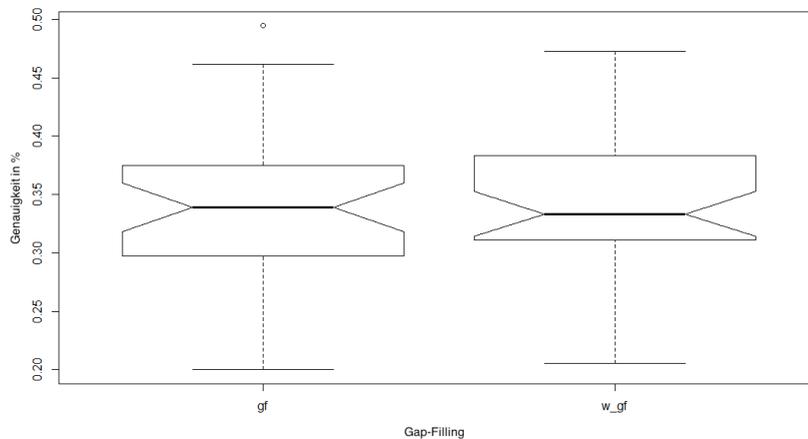
## 5.4 Kombinationen zwischen den drei Faktoren

### 5.4.1 Genauigkeit

Im Folgenden betrachten wir Kombinationen zwischen den beiden Faktoren Distanz und Länge in Abhängigkeit der Genauigkeit. Der entsprechende Boxplot ist in Abbildung 5.9 zu sehen. In diesem Boxplot kann man das gleiche Muster sehen wie in den oberen Genauigkeitplots: Bei konstanter Linklänge steigt die Genauigkeit mit steigender Distanz. Das gilt für jeden der vier Blöcke. Laut der statistischen Auswertung liegen signifikante Differenzen vor: ( $\chi^2 = 73.898; p = 2.205e - 11$ )

Die Kombination zwischen Distanz und Gap-Filling in Abhängigkeit der Genauigkeit ist in Abbildung 5.7 zu sehen. Je größer die Distanz der Achsen mit Gap-Filling, desto besser war die Genauigkeit. Zu erkennen ist, dass die Versuchspersonen bei der großen Distanz viel besser geschätzt haben, als bei der Variante mit Gap-Filling. Dies wird durch den Statistiktest bestätigt: ( $\chi^2 = 36.21; p = 0.0000008623$ ) Die Post-Hoc-Analyse zeigt, dass folgende Paare signifikante Differenzen aufweisen:  $gf_{50} - gf_{150}; gf_{50} - w_gf_{150}; gf_{50} - gf_{300}; w_gf_{50} - gf_{150}; w_gf_{50} - w_gf_{150}; w_gf_{50} - gf_{300}; gf_{150} - w_gf_{300}; w_gf_{150} - w_gf_{300}$

In Abbildung wird die Kombination von Linklänge und Gap-Filling gezeigt. Dieses Bild zeigt das gleiche Verhalten wie bei der Linklänge in Abhängigkeit zu Genauigkeit. Die Boxplots mit und ohne Gap-Filling sehen fast gleich aus. Laut der statistischen Auswertung liegen keine signifikante Differenzen vor.

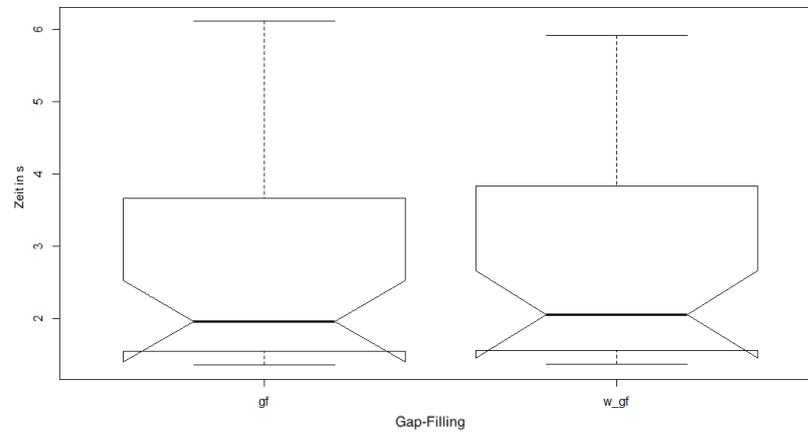


**Abbildung 5.5:** Genauigkeit in Abhängigkeit des Gap-Filling: Die horizontale Achse stellt das Gap-Filling und die vertikale Achse die Genauigkeit in Prozent dar

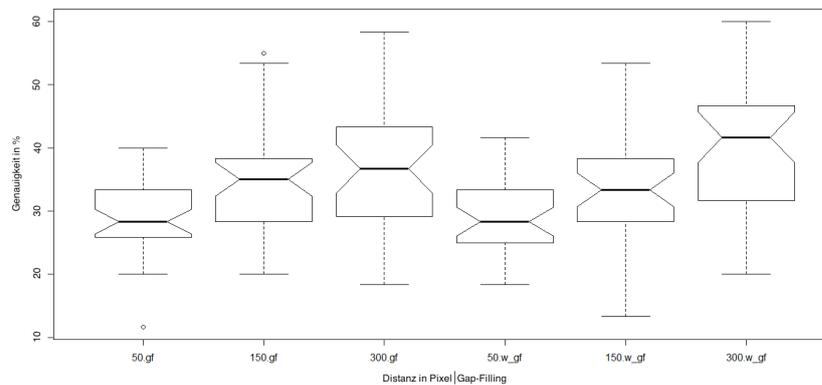
### 5.4.2 Zeit

Die Kombination Distanz und Gap-Filling in Abhängigkeit von Zeit ist in Abbildung 5.10 zu sehen. Hier sind keine Schwankungen zu sehen. Laut der statistischen Auswertung liegen keine signifikante Differenzen vor: ( $\chi^2 = 0.48665$ ;  $p = 0.9926$ )

Auch bei der Kombination Distanz und Linklänge ist nichts zu sehen, siehe Abbildung 5.12. Erneut liegen keine signifikanten Differenzen laut der statistischen Auswertung vor: ( $\chi^2 = 5.5765$ ;  $p = 0.9001$ )

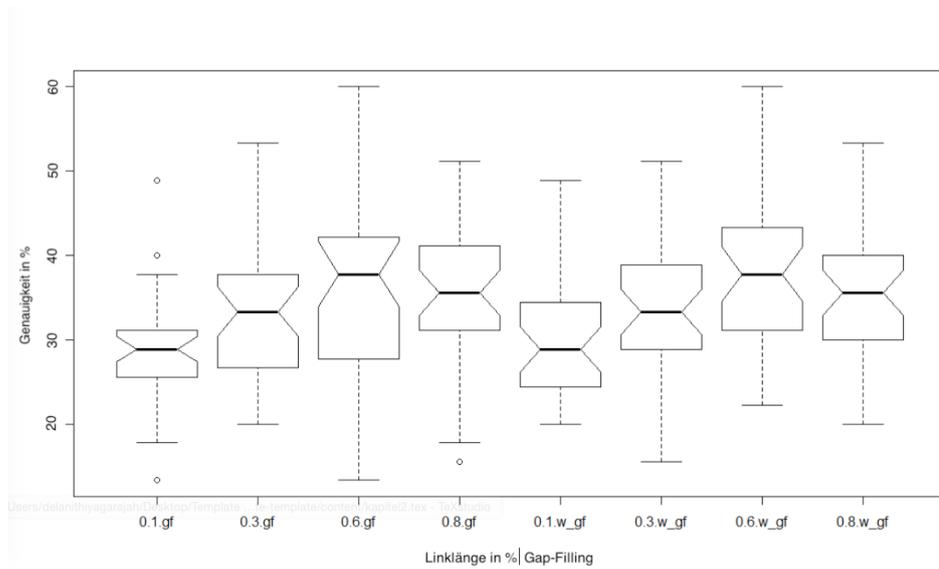


**Abbildung 5.6:** Zeit in Abhängigkeit des Gap-Filling: Die horizontale Achse stellt das Gap-Filling und die vertikale Achse die Zeit in Sekunden dar

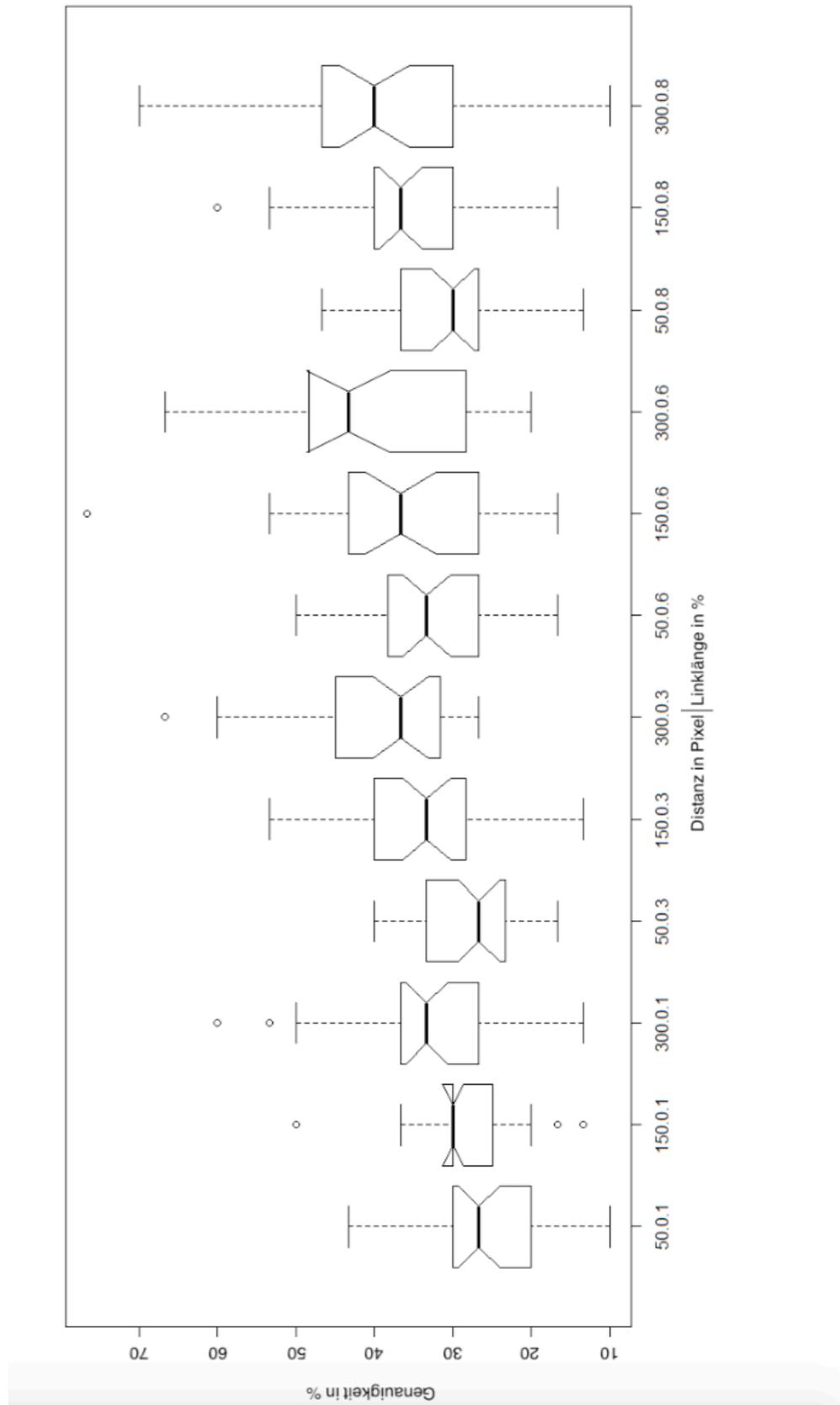


**Abbildung 5.7:** Genauigkeit in Abhängigkeit der Kombinationen Distanz und Gap-Filling: Die horizontale Achse stellt die Kombination von Distanz und GapFilling dar. Die vertikale Achse stellt die Genauigkeit in Prozent dar.

## 5.4 Kombinationen zwischen den drei Faktoren

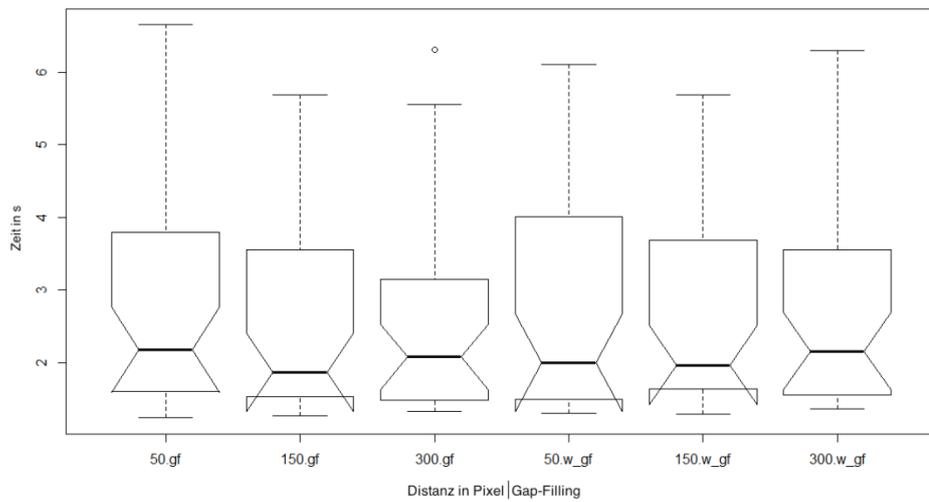


**Abbildung 5.8:** Genauigkeit in Abhängigkeit der Kombinationen Linklänge und Gap-Filling: Die horizontale Achse stellt die Kombination von Linklänge und GapFilling dar. Die vertikale Achse stellt die Genauigkeit in Prozent dar.

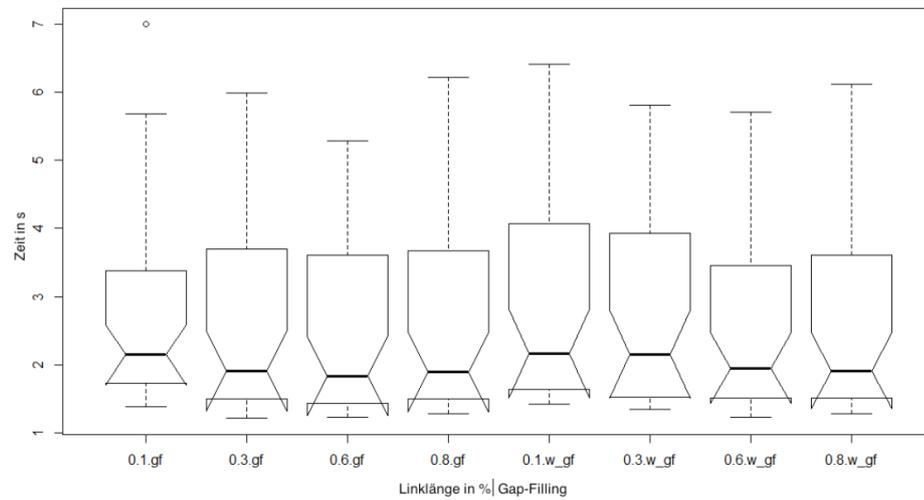


**Abbildung 5.9:** Genauigkeit in Abhängigkeit der Kombinationen Distanz und Linklänge: Die horizontale Achse stellt die Kombination von Distanz und Linklänge dar. Die vertikale Achse stellt die Genauigkeit in Prozent dar.

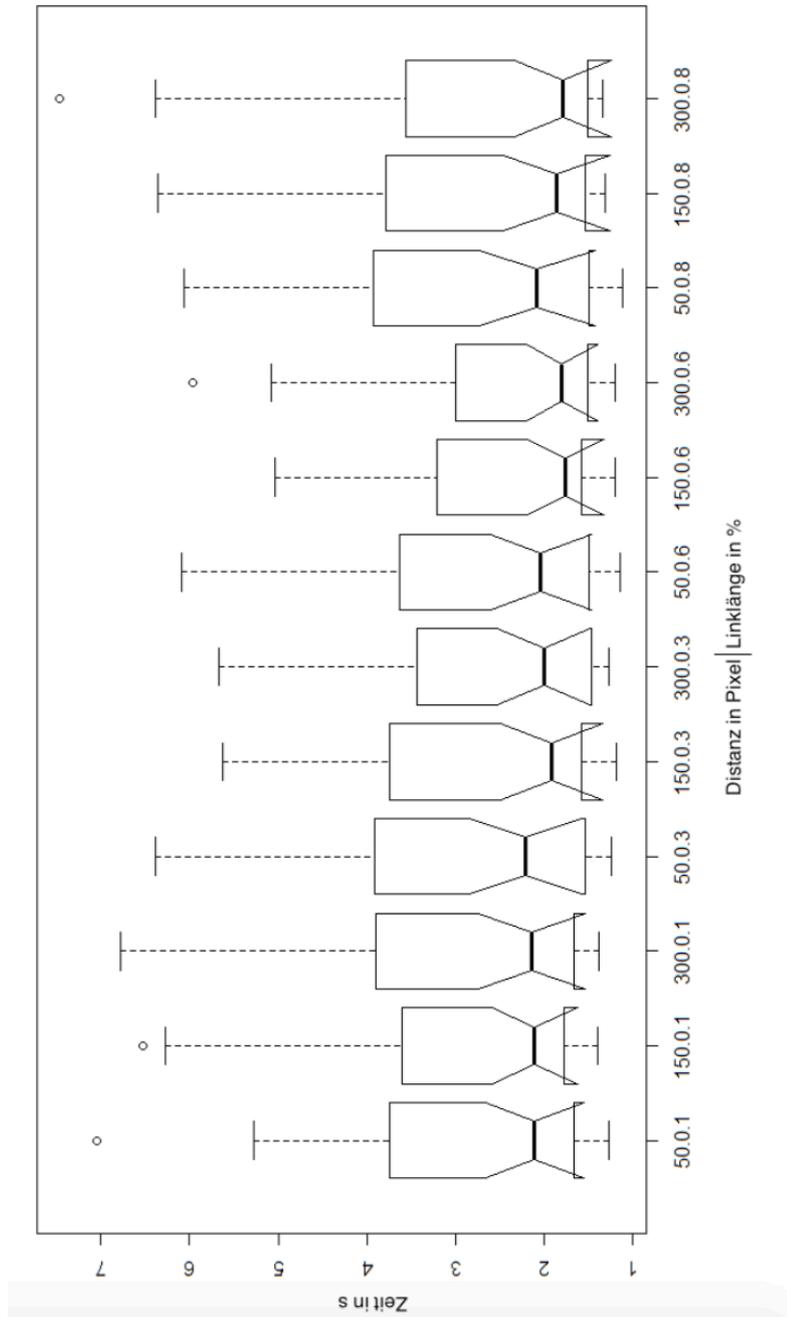
## 5.4 Kombinationen zwischen den drei Faktoren



**Abbildung 5.10:** Zeit in Abhängigkeit der Kombinationen Distanz und Gap-Filling: Die horizontale Achse stellt die Kombination von Distanz und GapFilling dar. Die vertikale Achse stellt die Zeit in Sekunden dar.



**Abbildung 5.11:** Zeit in Abhängigkeit der Kombinationen Linklänge und Gap-Filling: Die horizontale Achse stellt die Kombination von Linklänge und GapFilling dar. Die vertikale Achse stellt die Zeit in Sekunden dar.



**Abbildung 5.12:** Zeit in Abhängigkeit der Kombinationen Distanz und Linklänge: Die horizontale Achse stellt die Kombination von Distanz und Linklänge dar. Die vertikale Achse stellt die Zeit in Sekunden dar.



## 6 Zusammenfassung und Ausblick

Durch die Studie können wir schlussfolgern, dass die H1 teilweise bestätigt werden kann. Das heißt, geringere Linklänge führt zu geringer Genauigkeit. Die Linklänge hat jedoch keine Auswirkungen auf die Bearbeitungszeit. Die Hypothese 2 kann ebenso nur teilweise bestätigt werden. Die Distanz hat nur Auswirkungen auf die Genauigkeit. Die Hypothese 3 kann laut den Ergebnissen des Statistikttests nicht bestätigt werden. Die Resultate zeigen somit, dass es keine signifikanten Unterschiede bei der Bearbeitungszeit, aber bei der Genauigkeit in Bezug auf die Faktoren Linklänge, Gapfilling und Distanz geben.

### Ausblick

Wir haben ein Studie mit 36 Teilnehmern durchgeführt um die Auswirkungen der partiellen Kanten in parallelen Koordinaten auszuwerten. Die Studie zeigt, dass sich bei der Bearbeitungszeit der Aufgaben nichts ändert, sich jedoch die Fehlerraten erhöhen. Daraus kann man schließen, dass partielle Kanten die Fehlerraten erhöhen und somit die Reduzierung der visuellen Unordnung zum schlechten führt. Weiterhin sollten Studien mit partiellen Kanten durchgeführt werden um die Resultate wirklich zu bestätigen.



# Literaturverzeichnis

- [1] A. Engelhardt. „Streudiagramme“. In: (). URL: <http://www.crashkurs-statistik.de/streudiagramme/> (zitiert auf S. 16).
- [2] J.Heinrich, D.Weiskopf. „State of the Art of Parallel Coordinates“. In: *Eurographics 2013 . State of the Art Reports* (2013) (zitiert auf S. 11, 14).



## **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