

Projektbericht zum Modul Information Retrieval und
Visualisierung Sommersemester 2021

Bike Buyers 1000

Floyd Spuhler

10.09.2021

Inhaltsverzeichnis

1	Einleitung	3
1.1	Anwendungshintergrund	3
1.2	Zielgruppen	5
1.3	Überblick und Beiträge	6
2	Daten	6
2.1	Technische Bereitstellung der Daten	7
2.2	Datenvorverarbeitung	8
3	Visualisierungen	9
3.1	Analyse der Anwendungsaufgaben	9
3.2	Anforderungen an die Visualisierungen	9
3.3	Präsentation der Visualisierungen	9
3.3.1	Visualisierung Eins	10
3.3.2	Visualisierung Zwei	11
3.3.3	Visualisierung Drei	12
3.4	Interaktion	13
4	Implementierung	14
5	Anwendungsfälle	16
5.1	Anwendung Visualisierung Eins	16
5.2	Anwendung Visualisierung Zwei	17
5.3	Anwendung Visualisierung Drei	19
6	Verwandte Arbeiten	20
7	Zusammenfassung und Ausblick	21

1 Einleitung

Durch das wachsende Umweltbewusstsein in der Bevölkerung rückt besonders das Fahrrad als klimaneutrales Verkehrsmittel in den Fokus [1]. Bereits 2019 betrug der Gesamtbestand an Fahrrädern in Deutschland fast 76 Millionen [2, 3]. Die anhaltende Covid-19 Krise verstärkt den zunehmenden Fahrrad-Boom für die Freizeitgestaltung und den Pendelverkehr [4]. Gründe hierfür liegen beispielsweise bei der Sportmöglichkeit als Ausgleich zu geschlossenen Sportstätten, sowie im Individualverkehr für das Einhalten der Abstandregeln als Alternative zum Nahverkehr. Dabei wird sich häufig für das Fahrrad anstelle des Autos entschieden. [5, 6]. Passend dazu gaben im Rahmen einer Studie 61% von 782 Fahrradfahrenden an, dass Sie während der Covid-19 Krise bei Kurzstrecken auf das Fahrrad zurückgreifen, weswegen auch die Nachfrage nach dem Ausbau von Fahrradstrecken steigt. [7]. Um die Sicherheit von Fahrradfahrenden im Straßenverkehr zu verbessern, nimmt die Planung und der Bau neuer Fahrradstrecken in deutschen Städten zu [7, 8, 9]. Durch die Krise bedingte Lieferverzögerungen und Knappheiten führen zu einer hohen Nachfrage und steigenden Preisen. Auch wärmere Winter begünstigen eine längere Fahrradsaison und sorgen dadurch zusätzlich für Lieferengpässe [10]. Gleichzeitig werden Neuentwicklungen, wie E-Bikes, immer beliebter, was sich an der Gesamtabatzbeteiligung im Jahr 2020 von 38,7% in Deutschland bemerkbar macht [10]. In der Schweiz ist das Fahrrad nicht zuletzt durch den staatlich geförderten Ausbau von Fahrradspuren zum beliebtesten Verkehrsmittel geworden [4].

Lieferengpässe erschweren die Kapazitätsplanung und damit die Kundenbeziehung. Eine weitere Herausforderung ist der beschriebene Ausbau von Fahrradstrecken und die damit verbundene Planung. Daraus lassen sich die folgenden Fragen ableiten:

- Gibt es charakteristische Eigenschaften, die zum Fahrradkauf führen? Gibt es hierbei Zusammenhänge?
- Welche Anpassungen sind durch den Fahrrad-Boom in Bezug auf die Infrastruktur vorzunehmen?

Das Ziel dieser Arbeit ist es, diese Fragestellungen aufzugreifen und mit Hilfe von Visualisierungstechniken, wie dem Scatterplot, Parallelen Koordinaten und dem Baumdiagramm zu beantworten. Durch diese Darstellungen können Zusammenhänge hervorgehoben werden und Kernaussagen entnommen werden. Durch die Interaktionsmöglichkeiten mit den Modellen können eigene Erkenntnisse oder Trends gewonnen und individuell hervorgehoben werden.

1.1 Anwendungshintergrund

Diese Forschungsarbeit bereitet Informationen auf, die interessante Einblicke in das Kaufverhalten von Fahrradkaufenden geben. So lassen sich anhand von Einkommensdaten, dem Alter und dem Bildungshintergrund Kundengruppen ableiten, auf welche in allen Wertschöpfungsebenen eingegangen werden kann. Hersteller müssen beispielsweise die Rahmengröße auf das Alter abstimmen. Einkommensstarke Kunden setzen den Fokus auf hochwertige Materialien

(z.B. Carbon) und erwarten dabei technische Innovation, was die aktuelle Kooperation eines Fahrradunternehmens mit dem Premium-Autohersteller Porsche für ein E-Bike zeigt [11]. Diese werden häufig von einkommensstarken, älteren und männlichen Kunden gekauft [12, 13]. Damenfahrräder weisen schmalere Lenker- und nähere Bremsgriffe zur handlicheren Bedienung auf. Speziell Damenstadtfahrräder eignen sich auf Grund der besseren Abstiegsmöglichkeit auch für ältere oder versehrte Menschen im Stadtverkehr [14]. Personen die das Fahrrad zum Pendeln nutzen sind eher auf ein Stadt- als ein Mountainbike angewiesen. Darüber hinaus gibt es eine Vielzahl weiterer Verwendungsbereiche für Fahrräder, wie das Trekkingbike, Rennrad oder das Crossbike. Abgestimmt auf den vorliegenden Datensatz wird auf die Vorteile des CityBikes, insbesondere in urbanen Regionen mit Kurzstrecken eingegangen. Die übrigen Fahrradarten heben sich durch ihre Funktion für Sport, längere Touren oder unebenes Terrain hervor [15]. Durch die Datenaufbereitung der Arbeitswegdistanzen können Bauunternehmen in Innenstädten gezielter Fahrradwege bauen. Diese Anwendungsfelder werden über die drei verschiedenen Visualisierungsanwendungen aufgegriffen.

Mit einem Scatterplot, bestehend aus X und Y Achse, können Zusammenhänge von Y in Abhängigkeit von X festgestellt werden. Diese Zusammenhänge werden in Form von Punkten, die zwischen beiden Achsen liegen und einen Achsenwert darstellen, visualisiert. Dabei fallen die Beziehungen zwischen den beiden Punkten je nach Verwendungszweck unterschiedlich stark aus [16, 17, 18]. Der Scatterplot wird für die Daten zu Fahrradkunden als erste Darstellung von Zusammenhängen verwendet.

Mit der zweiten Visualisierung, Parallelen Koordinaten, können Zusammenhänge multidimensionaler Daten besser als über Punkte dargestellt werden. Punkte werden dabei über Linien und Achsen abgebildet [19]. Jede ausgewählte Variable wird durch eine eigene vertikale Achse symbolisiert und parallel zu den anderen Variablen nebeneinander angeordnet. Alle dazugehörigen Datenwerte werden über eine Linie miteinander verbunden [20]. Eine Linie, welche über alle Achsen abgebildet wird, heißt Polygonlinie [21].

Die vertikalen Achsen sind über Linien miteinander verbunden. Die Auf- und Ab-Bewegungen der Linien zeigen Werteveränderungen auf [22]. Dabei sind die Achsen parallel zueinander angeordnet. Parallele Koordinaten bieten einen guten Datenüberblick. Eine Gefahr stellt allerdings die Überlappung von Linien dar, wenn zu viele Daten verwendet werden [23].

Über ein explizites Baumdiagramm lassen sich Daten und deren Beziehungen untereinander anordnen, wodurch eine übersichtliche Datenstruktur entsteht und sich Daten schnell wiederfinden lassen. Die Baumstruktur besteht aus Knoten und Kanten. Zwei Knoten sind jeweils über eine Kante miteinander verbunden. In der Baumstruktur muss ein Knoten vorhanden sein, der keinen Vorgänger hat. Dieser Knoten wird Wurzel genannt. Dessen Folgeknoten werden Nachfolger genannt. Über die Wurzel führen nur azyklische Pfade und zu jedem Knoten nur ein Pfad. Durch die unterschiedlichen Pfade und Verzweigungen der unterschiedlichen Daten entsteht die Baumstruktur [24]. Diese Visualisierungstechnik eignet sich für die übersichtliche Darstellung bestimmter Daten zu Fahrradkäufen, da diese Daten Strukturen aufweisen, die sich zur Hierar-

chiedarstellung eignen. Im Datensatz sind des Weiteren Abfragen erhoben worden, ob sich eine Person ein Fahrrad gekauft hat oder nicht. Dadurch lässt sich der Datensatz als Erstes aufteilen. Des Weiteren sollen globale Unterschiede dargestellt werden, um auf territoriale Umfelder eingehen zu können. Ein weiterer Faktor ist die Pendeldistanz vom Wohnsitz der Befragten zur Arbeit, welche verschiedene Antwortmöglichkeiten zulässt. All diese Faktoren können kombiniert in die Baumhierarchie übertragen werden.

1.2 Zielgruppen

Dieser Forschungsbericht richtet sich vor allem an die Anbieterseite auf den B2C-Fahrradmarkt. Auf der Anbieterseite sind alle in der Lieferkette vorhandenen Unternehmensbrachen betroffen. Die Hersteller haben mit dem Materialmangel zu kämpfen [10]. Den Fahrradläden macht auf Grund durch Covid-19 geschlossener Geschäfte und günstigeren Preisen der Onlinehandel Konkurrenz [25, 26]. Bauunternehmen, die Fahrradspuren bauen, haben mit Rohstoffmangel ebenfalls Probleme [27]. Aus den beschriebenen Szenarien lassen sich drei Hauptzielgruppen, neben Fahrradinteressierten, herausfiltern, an welche sich dieser Visualisierungsbericht richtet.

- **Fahrradhersteller:**

Fahrradhersteller benötigen nicht nur wegen des Aspektes der Materialknappheit spezifische Informationen zu den personenbezogenen Merkmalen potenzieller Kunden, wie z.B. Alter und Einkommen. Dadurch können Materialien zielgerichteter bestellt und verwendet werden. Über das Alter lässt sich die Rahmengröße ableiten. Das Einkommen kann Informationen über das Interesse an hochwertigen Materialien und entsprechenden Qualitätsansprüchen liefern. Mit Hilfe dieser Informationen können zielgerichteter Fahrräder für die verschiedenen Verwendungszwecke, wie Mountainbike oder Stadtrad hergestellt werden.

- **Fahrradhandel:**

Für Fahrradverkäufer spielt vor allem der Verwendungszweck des potenziellen Kunden eine übergeordnete Rolle beim Fahrradkauf. Die Wahl des richtigen Modells unterscheidet sich für die Freizeit (Mountainbike) mit weiten Distanzen vom Gebrauch für die Stadt mit geringeren Distanzen (Citybike). Für weitere Distanzen eignen sich Mountainbikes besser als für die Fahrt in ebenerdigem Terrain, wie asphaltierten Straßen für Stadträder. Idealerweise betreibt diese Zielgruppe einen eigenen Onlineshop zum Fahrradvertrieb und benötigt Informationen für die zielgerichtete Kundengruppenwerbung.

- **Bauunternehmen mit dem Fokus auf Fahrradinfrastruktur:**

Auch für Unternehmen aus der Baubranche mit dem Fokus auf die Infrastruktur für Fahrradwege ist diese Arbeit eine geeignete Anlaufstelle für Informationen zum Einsatz des Fahrrads in Bezug auf den Arbeitsweg. Daten zu Pendelwegen müssen für diese Zielgruppe besonders aufgegliedert vorliegen, da Bauunternehmen somit Informationen über die

benötigten Distanzen neuer Fahrradwege erhalten und besonders in Städten nur begrenzt Raum zur Verfügung haben. Durch Visualisierungen der Fahrradkaufdaten wird diese Zielgruppe bei der Bauplanung unterstützt.

Dieser Visualisierungsbericht ermöglicht es den oben genannten Unternehmen eine bessere Kundenmarktsegmentierung zu betreiben. Kurzfristig können durch die Ergebnisse dieses Berichtes Ressourcen sparsam eingesetzt werden (v.a. Hersteller, Baubranche). Langfristig können besonders Fahrradgeschäfte von diesem Bericht profitieren, da sie durch die personenbezogenen Daten optimale Kundenakquise und Kundenberatung garantieren können und anhand von Einkommensparametern Preise bestmöglich bilden können.

1.3 Überblick und Beiträge

Die durch Kaggle bereitgestellten Daten bestehen aus demographischen Kundeninformationen, wie Alter, Geschlecht, Familienstand, Informationen zum finanziellen Hintergrund, wie Autoanzahl oder zum Immobilienbesitz und Hintergrundinformationen zu Bildungsstand und Berufsposition. Daraus geeignete Daten werden über die drei ausgewählten Visualisierungstechniken Scatterplot, Parallele Koordinaten und Baumhierarchie abgebildet, um den in 1.2 angesprochenen Zielgruppen eine visuelle Aufbereitung dieser Kundendaten zu vermitteln. Über den Scatterplot können jeweils zwei Dateneigenschaften, wie Alter und Einkommen der Personen, einander gegenübergestellt werden. Bei der Anwendung kann über Buttons selbst ausgewählt werden, welche beiden Eigenschaften angezeigt werden sollen. Darüber hinaus sind die Punkte verschieden eingefärbt, wodurch eine weitere intuitive Unterscheidung ermöglicht wird. Mit den parallelen Koordinaten haben Interessenten die Möglichkeit über vertikale Achsen Werte miteinander zu vergleichen. Buttons ermöglichen hierbei die dynamische Achsenverschiebung. Die Baumhierarchie lässt die Daten anhand wesentlicher Eigenschaften in verschiedenen Ebenen darstellen. Darüber hinaus wurden Buttons integriert mit denen zwischen den einzelnen Visualisierungen und einer Startseite gewechselt werden kann.

2 Daten

Die diesem Projektbericht zugrundeliegenden Rohdaten entstammen einem Datensatz des „Kaggle“-Accounts von Heeral Dedhia [28], welche Antworten von 1.000 befragten Personen zum Thema Fahrradkauf in der zum Zeitpunkt der Arbeit zweiten Version bereitstellt. Die Nutzerin hat diese Daten zuletzt im Jahresverlauf 2020 dahingehend angepasst, dass fehlende Werte ergänzt und Ausreißer konfiguriert wurden. Ein genaues Datum sowie spezifische Hintergrundinformationen zur Erhebungsform sind hierbei unbekannt. In dieser aktuellsten Version liegen 13 verschiedene Attribute zu den 1000 befragten Personen vor. Die Nutzerin hat zwei verschiedene Datensätze bereitgestellt, die sich durch fehlende Werte voneinander unterscheiden. Um bei einer Datenvorverarbeitung keine Daten zu vergessen und die Funktionsfähigkeit des Elm CSV-Decoders zu

gewährleisten, stellt die bereinigte Datei „bikebuyersclean.csv“ die Grundlage für dieses Visualisierungsprojekt dar.

Zu allen befragten Personen wurde eine eindeutige „ID“ vergeben, welche ein Int-Typ ist. Auf diese ID wurde im Rahmen der Baumdarstellung zurückgegriffen. Die nächsten beiden Spalten „Marital Status“ , „Gender“ und „Children“ und „Age“ als Int geben als String-Datentyp Aufschluss über den sozialen Familienstand, Geschlecht, sowie Anzahl der Kinder und Alter der befragten Person. Die Spalten „Income“, „Education“ , „Occupation“ geben Aufschluss über die berufliche Karriere. In Verbindung mit den Spalten „HomeOwner“ und „Cars“ lässt sich der Wohlstand der Person aus Immobilienbesitz und Autoanzahl interpretieren. Das Attribut „Commute Distance“ gibt Aufschluss über die Distanz zur Arbeit dient der Befragung zur Entfernung zwischen Wohnort und Arbeitsstätte. Die angegebene Entfernungen wurden in fünf Kategorien zwischen Null und über zehn Meilen eingeteilt. Die Spalte „purchasedBike“ erfragt den Fahrradkauf, wodurch sich die Personen in zwei Gruppen aufteilen. Zusätzlich wurde zu jeder befragten Person durch die Spalte „region“ eine Region für den Wohnort zugeordnet.

Die demographischen Personendaten in Verbindung mit dem Einkommen eignen sich besonders für die Anforderungen an die Fahrradersteller. Für den Fahrradhandel sind diese Informationen, verknüpft mit weiteren Daten zum Wohlstand , wie Immobilienbesitz, hilfreich. Für die angesprochene Baubranche sind vor allem die Kategorien zur Pendeldistanz für die Infrastrukturprojekte von Bedeutung. Insgesamt hat jede der Zielgruppen die Möglichkeit sich individuell auf eine Kundengruppe oder einen Datenzusammenhang zu spezialisieren, was den Wettbewerb vor allem unter den Fahrradunternehmen belebt. Somit bietet der Datensatz den nötigen Mehrwert, um den unterschiedlichen Anforderungen der drei Zielgruppen zu entsprechen und die eingangs aufgestellten Fragen zu beantworten.

Um eine geeignete Überblicksmöglichkeit über diese potenziellen Kundengruppen zu schaffen, musste der dafür notwendige Datensatz für die Baumhierarchie angepasst werden.

2.1 Technische Bereitstellung der Daten

Die dem Kaggle Account [28] entstammenden Rohdaten in der Datei „bike buyers clean.csv“ wurden in das Github Repository des Autors übertragen, damit die im Laufe der Projektarbeit eventuell nötigen Verarbeitungsschritte ersichtlich wären. Die ist im Ordner „Daten zum Laden“ abgelegt. Für die beiden Elm Dateien „Scatterplot.elm“ und „Parallele Koordinaten.elm“ wurden die vollständigen Daten der Datei „bike buyers clean.csv“ als gemeinsamer Csv-String in die jeweiligen ELM Dateien geladen. Dieses Vorgehen hat den Vorteil, dass die beiden Anwendungen unabhängig von Serverproblemen oder Link-Veränderungen zur Verfügung stehen. Das für die Baumhierarchie notwendige JSON-Format wird im Ordner „JSON“ durch die Datei „DatenvorverarbeitungohneCarWolrdwide.json“ bereitgestellt und über einen Link in die Elm-Datei „Baumhierarchie.elm“ geladen, wo diese vor-verarbeiten Daten aufgegriffen werden.

2.2 Datenvorverarbeitung

Bei der Entscheidung, für die Visualisierungen Eins „Scatterplot.elm“ und Zwei „ParalleleKoordinaten.elm“ den CSV-Datensatz als String in die beiden Programmcodes zu kopieren und über einen CSV-Decoder zu entpacken, war keine Datenvorverarbeitung notwendig. Die Daten werden fehlerfrei durch den Decoder gelesen.

Für die dritte Visualisierung war eine Datenbearbeitung vor der Verwendung in der dazugehörigen ELM-Datei „Baumhierarchie.elm“ notwendig. Dafür wurde das Datenverarbeitungsprogramm Excel von Microsoft verwendet. Im ersten Schritt wurde der CSV-String bestehend aus allem Attributen in einer Tabellenspalte in die 13 einzelnen Spalten transformiert. Danach wurden alle für die Baumdarstellung nicht benötigten Attribute gelöscht, sodass „ID“, „cars“, „Region“ und „purchasedBike“ übrig geblieben sind. Dieser Entwicklungsstand kann der Datei „Visualisierung3 Vorverarbeitung.xlsx“ entnommen werden. Im weiteren Entwicklungsverlauf wurde sich dazu entschieden, alle Zeilen auszuklammern, die ein oder mehrere Autos besitzen. Dadurch ist das Ergebnis für den Vergleich der Pendeldistanzen zwischen Fahrradkaufenden und Nichtkaufenden besser interpretierbar und der auftretende Effekt stärker. Parallel dazu wurde das Länderbeispiel aus der Übung zehn als Grundlage verwendet, um ein Gefühl für den JSON-Syntax zu gewinnen. Die komplette Datei wurde entkernt, das heißt, die frühere Datenstruktur nach Regionen und Ländern wurde entfernt. In das so bereinigte JSON Grundgerüst wurde die Gabelung Fahrradkauf (Ja/Nein), die Verzweigung in die drei Regionen pro Seite und anschließend die letzte Unterteilung mit den Distanzen und darunter die IDs aufgebaut. Somit können unter der Kategorisierung der Pendeldistanzen alle dazugehörigen IDs aufgeführt werden, was Anwendenden direkt relevante Zusammenhänge präsentiert. Für diese letzte Gabelung wurde zur Zeitersparnis und Fehlervermeidung, statt manueller Bearbeitung in VSC, Excel verwendet, um mit dem Befehl

```
=VERKETTEN("{\"data\": { \"id\":\"\"; A57;\"\" }},")
```

Abbildung 1: Verwendeter Excelbefehl für schnelle Datenvorverarbeitung, Quelle: Eigene Darstellung

den letzten Wurzelstrang in der JSON-Notation aufzubauen. „A57“ steht beispielhaft für eine ID-Zelle und wurde auf alle Zeilen übertragen. Dadurch wird der letzte benötigte Strang des Baumdiagramms vom ELM JSON-Decoder erkannt. Dieser letzte Vorverarbeitungsschritt in Excel kann der Datei „BikeBuyers Vorverarbeitung für JSON“ entnommen werden. Die fertige Vorverarbeitung der im ELM Programm verlinkten JSON-Datei ist im Ordner „JSON“ unter „DatenvorverarbeitungohneCarWorldwide.json“ abrufbar.

3 Visualisierungen

In diesem Kapitel werden drei ausgewählte Visualisierungstechniken vorgestellt und diskutiert.

3.1 Analyse der Anwendungsaufgaben

Über den Scatterplot erhält besonders die Zielgruppen der Fahrradhersteller und des Fahrradhandels wichtige Informationen auf einen Blick, die bei der Produktion oder dem Fahrradverkauf unterstützen können. Mit den Parallelen Koordinaten können die Zusammenhänge der Zahlenwerte aus dem Datensatz besser untersucht werden. Die Röntgendarstellung ermöglicht es, Überschneidungen gut zu erkennen. Hierdurch haben vor allem Fahrradhandelsgeschäfte (auch mit Onlineshop) die Möglichkeit Ihren Kundengruppen zielgerichteter Fahrräder anzubieten und Fahrradhersteller können die notwendigen Fahrradkomponenten und Materialien besser auf die Kundengruppen abstimmen. Durch das genaue Eingehen auf die Kundenbedürfnisse sind mit dem Fahrradkauf positive Assoziationen verbunden, wodurch auch die Value Proposition für die Kundengruppen gestärkt wird.

Durch die Visualisierung von Hierarchien über das Baumdiagramm erhalten Bauunternehmen auf einen Blick wesentliche Informationen über Entfernungen für die Infrastruktur, die diese und alle mit dem Bauvorhaben involvierten Gruppen in dieser Phase unterstützt.

3.2 Anforderungen an die Visualisierungen

Im ersten Kapitel wurde die eingehende Motivation beschrieben, den verschiedenen Zielgruppen bestmögliche Anhaltspunkte zu finden, um die Fahrradkundengruppe nachhaltig an die jeweilige Unternehmensbranche zu binden. Die ausgewählten Designs müssen die angesprochene Übersichtlichkeit, vor allem für die Ableitung von Infrastrukturmaßnahmen, über erhobene Parameter einhalten. Des Weiteren liegt ihr Wertversprechen für die Zielgruppen im Aufzeigen von Zusammenhängen, welche zu Fahrradkäufen führen und der Komplexitätsreduktion der Datenmenge. Die Herausforderung besteht darin, eine Balance zwischen wahrnehmbarer Übersicht der 1000 Daten zu Fahrradkäufen und genügend Mehrwert für die Anwendenden zu finden. Hierzu sollten die ausgewählten Attribute voneinander unterscheidbar sein um die Heterogenität der Darstellungen zu gewährleisten.

Des weiteren sollen Anwendende Interaktionsspielraum mit den Anwendungen haben, um für sie spezifischere Informationen entnehmen zu können. Anwendende müssen die Gewissheit haben, dass keine Datenwerte durch die Visualisierung verändert wurden und den Erhebungsdaten zu den Fahrradkäufen entsprechen.

3.3 Präsentation der Visualisierungen

In diesem Abschnitt werden die drei dem Projektbericht zugrunde liegenden Visualisierungen, der Scatterplot, Parallele Koordinaten und die Baumhierarchie zu den Fahrraddaten präsentiert.

3.3.1 Visualisierung Eins

Für die erste Darstellung wurde der Scatterplot gewählt. Dessen einfache Grundfunktionen bieten eine gute Einleitung in die Visualisierung der Fahrradwerte. Die in der Einleitung beschriebenen Merkmale, wie die Gegenüberstellung der X und Y Achse erfolgt über die Datenwerte Alter auf der X Achse und in dieser Abbildung über Einkommen. Anwendende haben darüber hinaus die Möglichkeit, über die oben gelegenen Buttons die Y-Achse des Scatterplots dynamisch anzupassen. Dabei können weitere Zahlenwerte des Fahrraddatensatzes wie Einkommen, die Kinder- und Autoanzahl in Abhängigkeit des Alters dargestellt werden und Die Punkte stellen jeweils einen Koordinatenpunkt bestehend aus Alter und Einkommen dar. Durch unterschiedliche Punkteinfärbung können weitere Unterscheidungen und Erkennungsmuster identifiziert werden. Um die Zielgruppen mit Zusatzinformationen zu versorgen, können beim Hovern mit der Maus über die Punkte neben dem exakten Einkommen auch Informationen ob ein Fahrradkauf getätigt wurde, welcher Berufsgruppe die befragte Person angehört und zum Immobilienbesitz eingeholt werden. Da für das Koordinatensystem keine Textwerte aufgeführt werden können, ermöglicht dieser Scatterplot wichtige Informationen, die keine Zahlenwerte sind mit in die Darstellung einfließen zu lassen.



Abbildung 2: Scatterplot für Fahrradkäufe, Quelle: Eigene Darstellung

Die in der Einleitung herausgearbeiteten Scatterplot-Anforderungen können, wie in Abbildung 2 entnehmbar, umgesetzt werden. Die beliebige Y-Achsenveränderung (oben links: Einkommen, Kinder, Autos) erweitert die Anforderungen dahingehend, dass Anwendende den Darstellungsinhalt und somit bestimmte Schwerpunkte selbst bestimmen können. Blaue Punkte symboli-

sieren, dass die befragte Person ein Fahrrad besitzt, während gelbe Punkte aussagen, dass die Person kein Fahrrad besitzt. Stärker sichtbare Punkte und jeweils kräftigere Blau- und Gelbfarben bedeuten, dass die gemachten Angaben in diesen Punkten miteinander übereinstimmen. Punkte mit jeweils schwächeren Blau- und Gelbtönen zeigen in dieser Koordinatenumgebung weniger Übereinstimmung. Farbmischungen, wie die hellgrün eingefärbten Punkte, zeigen hingegen, dass hierbei sowohl Fahrradkäufe wie auch Nichtkäufe vorliegen. Der Punkt, über den gerade gehovert wird, ist für diese Aktion in der Farbe rot dargestellt. Die damit zusammenhängende rote Anzeige liefert die Zusatzinformationen zum Punkt. Eine Alternative zum Scatterplot stellen die Zeitreihendiagramme dar, wodurch Zeitliche Verläufe anhand vorbestimmter Faktoren als Linien dargestellt werden. Da der vorliegende Datensatz keine zeitlichen Ausprägungen hat, welche chronologisch dargestellt werden könnten, sondern zeitlich unabhängige Werte beinhaltet wurde der Scatterplot als zweidimensionale Datenvisualisierung gewählt.

3.3.2 Visualisierung Zwei

Mit der zweiten Visualisierung, den Parallelen Koordinaten, können im Gegensatz zum Scatterplot mehrdimensionale Zahlenwerte gleichzeitig, über vertikale Achsen ohne X Achse mit Polygonlinien als Punktrepräsentation dargestellt werden. Für die Daten zu den Fahrradkäufen ergibt sich hierdurch den Vorteil, gleichzeitig alle relevanten Zahlenwerte in einer Visualisierung darzustellen, was sich in Abbildung 2 bemerkbar macht. Ebenso wie beim Scatterplot haben Anwendende die Möglichkeit die Darstellung über Buttons zu verändern. Durch die Buttons können die Achsen beliebig miteinander vertauscht werden und neue Zusammenhänge, je nach Betrachtungsziel, identifiziert werden.

Die Abbildung 3 zeigt von links nach rechts den Verlauf eines mehrdimensionalen Punktes beginnend mit dem Einkommen, der Kinder- und Autoanzahl und dem Alter der befragten Person. Der Vorteil im Vergleich zum Scatterplot ist, dass mehrere Zusammenhänge und Auffälligkeiten gleichzeitig erkennbar sind. Im Direktvergleich mit der Darstellung 2 gibt es keine Hoverfunktion, wodurch die Linien individuell nach verfolgbar wären. Durch die Wahl des dunklen Hintergrunds und der hellen Linienfarbe ergibt sich allerdings eine Röntgendarstellung. Durch die Linienüberlappung werden Identifikationsmuster sichtbar, was sich durch die kräftigeren Weißtöne erkennen lässt. Darüber hinaus bietet die durchgängige Achsenbeschriftung gute Anhaltspunkte für den Linienverlauf. Das Vorfiltern nach Region und Fahrradkauf im ELM-Programm verhindert eine „Überdarstellung“ und Reizüberflutung für die Anwendenden. Somit wurden nicht nur alle Anforderungen gemäß der Theorie richtig abgeleitet, sondern auch die Gefahr der „Übervisualisierung“ erkannt und verhindert. Neben Parallelen Koordinaten gibt es weitere Visualisierungsmöglichkeiten mehrdimensionaler Daten. Da Scatterplots, am besten für zweidimensionale Daten geeignet, bereits verwendet wurden, stellen Projektion und Sektion, Sternkoordinaten, K-Means und Datentinte eine Alternative im Rahmen der Vorlesung zu Parallelen Koordinaten dar. Mit Projektionen und Sektionen soll der mehrdimensionale Raum abgebildet werden. Eine hierfür anfallende hohe Anzahl an benötigten Darstellungen kommt in Bezug auf die Übersichtlichkeit

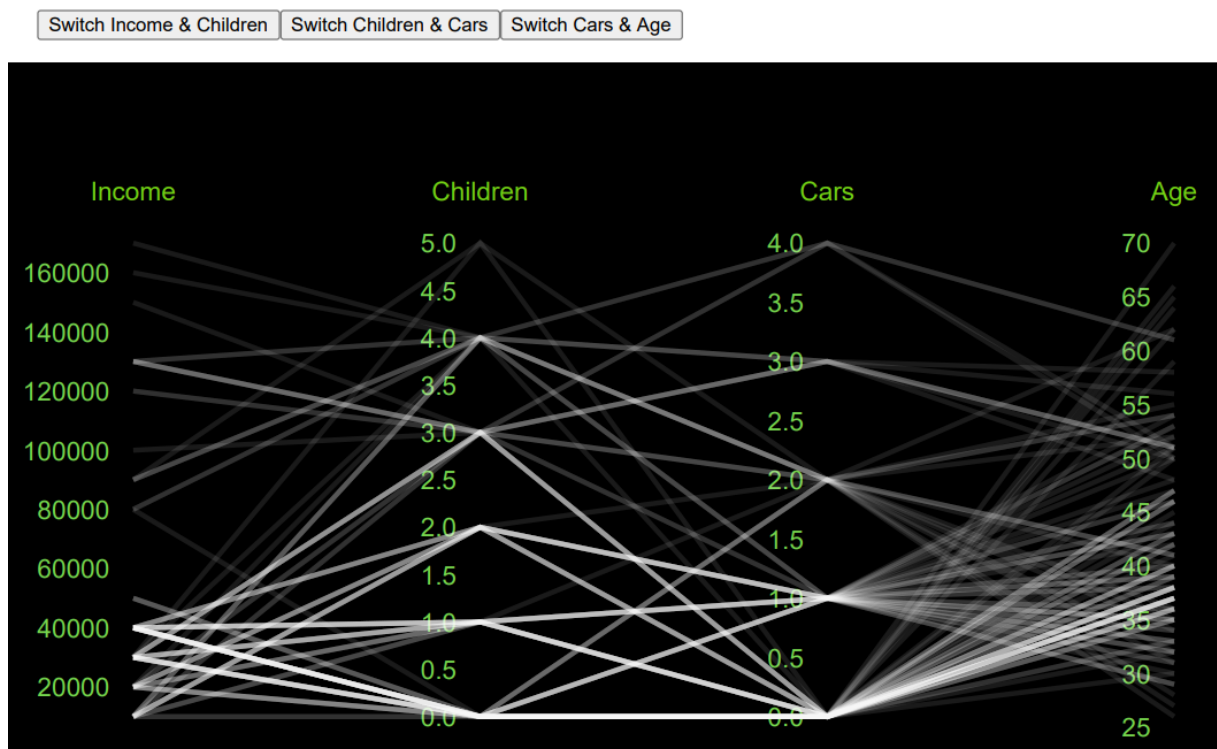


Abbildung 3: Parallele Koordinaten für die Fahrraddaten, Quelle: Eigene Darstellung

und Anschaulichkeit für die Zielgruppen nicht in Frage. Mit Sternenkoordinaten können mehrdimensionale Daten in 2D oder in 3D abgebildet werden. Der Name ist charakteristisch für die Achsenanordnung. Auf Grund der ungewöhnlichen Erscheinung und der besseren intuitiven Nachvollziehbarkeit der Fahrradkaufdaten wurde auf die Parallelen Koordinaten zurückgegriffen. Die Visualisierungstechniken K-Means und Datentinte fokussieren die Visualisierung durch Vorabberechnungen auf wenige Kerngedanken und verhindern somit einen Überblick über eine breitere Sichtweise, welche für alle Zielgruppen anvisiert wurde. Aus diesen Gründen wurde Visualisierung mehrdimensionaler Daten durch die Parallele Koordinaten präferiert.

3.3.3 Visualisierung Drei

Als dritte Visualisierungstechnik wurde die Baumhierarchie ausgewählt. Diese stellt ein klassisches Baumdiagramm, wie in der Theorie beschrieben dar, durch welches die Fahrradkäuferdaten in eine Hierarchie gebracht werden. Die Knoten stellen hierbei Entscheidungsmöglichkeiten dar, welche durch die Linien miteinander verbunden werden. Die erste Unterscheidung besteht beim Fahrradkauf. Hierbei werden die Angaben in Ja oder Nein unterteilt. Anschließend erfolgt die Unterteilung nach den Regionen der befragten Personen. Die letzte Verzweigung stellt die Pendeldistanz vom Wohnort zur Arbeit dar.

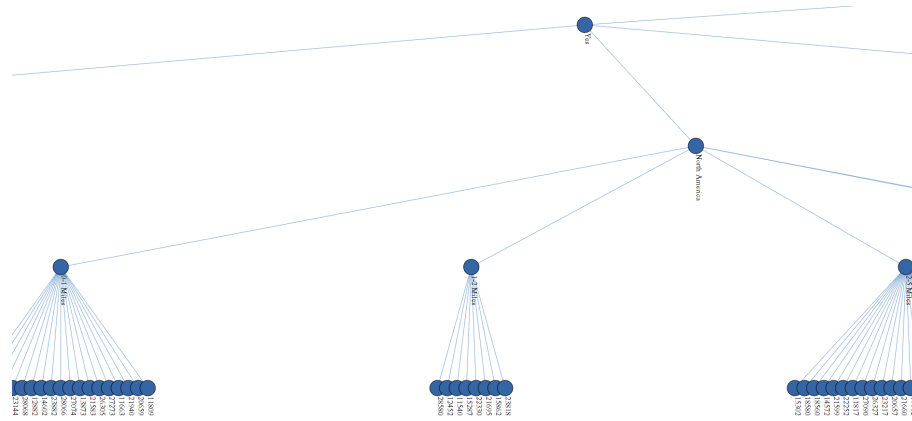


Abbildung 4: Ausschnitt aus der Baumhierarchie für die Fahrraddaten, Quelle: Eigene Darstellung

Die Abbildung 4 zeigt einen Ausschnitt aus der umfangreichen Baumdarstellung. Hierbei lassen sich die eingefärbten Knoten gut unterscheiden. Im Vergleich zu den beiden vorangegangenen Visualisierungen wurde auf eine Interaktionsmöglichkeit verzichtet. Die angeleiteten Baumhierarchie-Bedingungen, nämlich die übersichtliche Darstellung und Einteilung in Gruppen konnte umgesetzt werden. Trotz Vorfilterung der Parameter, dass die befragten Personen keine Autos besitzen, ist die Baumstruktur sehr umfangreich, weshalb auf weitere Unterteilungen, wie nach Familienstand, Beruf oder Bildungsstand verzichtet wurde, welche teilweise durch die Hoverfunktion im Scatterplot ersichtlich werden. Neben der Möglichkeit die dritte Visualisierung als Baum darzustellen, gibt es weitere Kategorien, welche allerdings die Darstellung komplexer machen. Hierunter fallen verschiedene Graphenanwendungen, die Mehrfachzuweisungen auf Knoten ermöglichen. Auf Grund der Zielsetzung, bei der Infrastrukturplanung visuell intuitiv zu unterstützen und aufzuzeigen, welche Pendeldistanzen auftreten, wurde auf eine komplexere Darstellungsmethode verzichtet.

3.4 Interaktion

Für die Visualisierungen Scatterplot und Parallele Koordinaten wurden zwei verschiedene Interaktionsmöglichkeiten in den jeweils zugrunde liegenden Code integriert.

Bei der Visualisierung des Scatterplots haben Anwendende die Option, die Gegenüberstellung des Alters auf der X Achse mit einem beliebig verfügbaren Zahlenwert des Bike-Buyers-Datensatzes auf der Y-Achse über das Klicken auf die verschiedenen Buttons anzeigen zu lassen. Dadurch kann der Fokus individuell auf bestimmte Sachverhalte gelegt werden, was insbesondere für Fahrradhersteller und Fahrradhändler interessant ist. Um die eingeschränkte Funktionalität der zweidimensionalen Zahlengegenüberstellung zu erweitern, wurde die Hover-Funktion integriert. Durch das Hovern über den Punkten des aufgespannten Koordinatensystems werden auch Textdaten als Informationen, wie der Beruf oder Immobilienbesitz ausgegeben. Daneben lassen

sich die exakten Datenwerte beim Einkommen anzeigen, welche lediglich als Achseneinteilung zu ungenau wären.

Die zweite Visualisierung, Parallele Koordinaten, bietet Interessenten die Möglichkeit über Buttons die Achsen nach belieben zu verschieben. Je nach Einstellung lassen sich Zusammenhänge der Polygonlinie hervorheben, was durch eine eindeutigere Überlappung der hellen Linien sichtbar wird. Anwendende werden dadurch ermutigt, selbst die Daten in eine für sie interessante Sichtweise zu verschieben, was beispielsweise bei Verkaufstrainings für den Fahrradhandel, aber auch für zielgerichtete Kundenwerbung hilfreich ist.

Für das Baumdiagramm wurde auf eine Interaktionsmöglichkeit verzichtet, da das Visualisierungsziel über eine Version erfüllt ist.

Über die Startseite „index.html“ kann aus dem eigenen, öffentlichen Github Repository eine Website dargestellt werden. Dabei wurden die drei Visualisierungen als HTML-Seite, hervorgehoben durch den Terminalbefehl in VSC „elm make“, jeweils verlinkt. Untereinander sind die HTML-Seiten miteinander vernetzt, sodass eine Interaktion möglich ist. Dabei können Anwendende, je nach Anwendungsziel, beispielsweise einen Zusammenhang im Scatterplot auf mehrdimensionaler Ebene mit Hilfe der Parallelen Koordinaten oder andersrum untersuchen, indem sie auf dafür konfigurierte Buttons zwischen den drei Visualisierungen und der Startseite wechseln können.

4 Implementierung

Die Ausgangsbasis für die Funktionsweise der Visualisierungen auf Codeebene stellen die Programmcodes mit ihren Funktionen aus der Übung dar. Für die Scatterplot-Visualisierung wurden die im Rahmen der Übungen eins bis drei entwickelten Codegerüste als Grundlage verwendet. Für die Parallelen Koordinaten ist die Übung 7 der Ausgangspunkt. Bei der Baumhierarchie galt die Übung 10 als Orientierung. Alle drei Visualisierungsprogramme beginnen mit dem Importieren der für die jeweiligen Elm-Programme notwendigen Pakete. Es wurden keine eigenen Module eingesetzt. Ein Überblick über alle verwendeten Pakete ist in der Readme.md im Github-Repository einsehbar.

Im Scatterplot-Code folgen als nächstes Funktionen welche die Darstellung für die spätere Scatterplot-Funktion, wie die Achsen oder die Größe, konfigurieren. Direkt im Anschluss folgen "typeFestlegungen zur Weiterverwendung in Funktionen der Elm-Datenstruktur, welche die Y-Achsenwerte definieren. Die folgende main-Funktion zeigt die Codegliederung der Elm-Datenstruktur auf. Die initFunktion legt die erste dynamische Scatterplot-Variante aus Alter und Einkommen fest. Die updateFunktion bestimmt den Darstellungswechsel über die Y-Achse. Als Teil der Elm-Struktur werden Subscriptions für die Main-Funktion gebraucht. Mit "view" wird die HTML-Anzeige für die Startdarstellung, die Buttoninteraktion und dadurch ausgelöstem Achsenwechsel bestimmt. Die nächsten drei Funktionen dienen der Visualisierungsspezifikation der Scatterplot-Punkte und der Scatterplot-Funktion. Darunter befinden sich die Funktionen,

welche die Möglichkeit zulassen sollen, den Datensatz beliebig zu skalieren und auch leere Werte im Programm zu filtern. Als letztes befinden sich Funktionen für das Decoden des großen CSV-Strings, welcher sich am Ende der Datei befindet. Diese letzten beiden Gliederungsfunktionen sind identisch mit dem Programm für die Parallelen Koordinaten.

Im Programm für die parallelen Koordinaten finden sich der gleiche Aufbau wieder. Nach den Paketimporten, der Darstellungskonfiguration der parallelen Koordinaten und den type-Funktionen zur Verwendung in der ELM-Datenstruktur wird im init die Startvisualisierung mit der Achsenreihenfolge festgelegt. Die beiden Funktionen indexSwap und update sind für den Achsenwechsel und die dafür notwendige Festlegung der Reihenfolge verantwortlich. Im let Block der view-Funktion werden zwei Parameter der Fahrraddaten vorgefiltert, um die Polygonlinien sichtbar darzustellen. Die view Funktion verwaltet die auf der HTML Seite dargestellten Inhalte. Über die parallelCoordinatesPlot-Funktion wird die Abbildung der vertikalen Achsen organisiert und über die Funktion colordescr die Achsenfarbe und Darstellung angepasst. Die letzten Funktionen dienen dem Herausfiltern potenziell zukünftig fehlender Werte und der Csv-String-Decodierung. Bei der Baumhierarchie wird nach dem Paketimport, der type-Festlegung und dem Konfigurieren der Elm-Datenstruktur über main, im init der vorbereitete Link zur JOSN Datei in das Programm geladen. Das Update dient zur korrekten Linküberprüfung und würde im Fehlerfall eine Benachrichtigung anzeigen. Danach folgen Funktionen zur Festlegen der Knoten und Kantendarstellung, sowie Layout-Konfigurationen. Nach dem view für die HTML-Anzeige unter Berücksichtigung der Konfigurationsfunktionen befindet sich der JSON-Tree-Decoder, der die verlinkte JSON-Datei entpackt.

Für die ersten beiden Visualisierungen wird ein CSV-Decoder benötigt. Während der Projekt-Konzeptionsphase wurde versucht, mit Hilfe des CSV-Decoders nach dem Vorbild aus Übung acht die Daten in das ELM-Programm zu laden. Allerdings ist der unübersichtliche Aufbau des Decoders nicht für die Projektarbeit mit den Fahrraddaten geeignet. Hierzu sind weitere Decoder-Funktionen notwendig, die zur Komplexität und Unübersichtlichkeit der Elm-Codes beitragen. Befürchtungen bei der Verwendung eines Decoders aus externen Links und mit der Projektarbeit verbundenen Dateiänderungen, sowie die in Kapitel 2.1 beschriebenen Vorteile waren ausschlaggebend für eine Alternativenrecherche zu intuitiv anwendbaren CSV-String-Decodern in der ELM-Dokumentation und auf GitHub. Ein Vorbild stellten hierzu die IRV-Car-Übungen mit den darin gespeicherten Daten dar. Der Unterschied ist, dass diese Car-Daten elmgerecht aufbereitet wurden, wodurch eigene Datentypen über "type" problemlos zum Filtern festlegbar sind. Zusätzlich sind die CSV-Daten in ELM nicht ohne Zusatzfunktionen einlesbar. Der Entwickler Brian Hicks bietet einen für die Anforderungen an die Fahrraddaten passenden Decoder. Dessen Paket für Elm wurde in die beiden Programme importiert und die Decoder-Funktion aufgebaut. Die Herausforderung in dieser Phase war die Implementierung des Decoders, sodass das Elm-Programm die Fahrraddaten richtig laden und darstellen kann. Nach erfolgreichen Tests dieser Anforderung wurde mit der zweiten Phase begonnen.

Diese Phase umfasst die Verbindung der decodierten Fahrraddaten mit den Codegerüsten aus

der Übung zum Scatterplot und zu den Parallelen Koordinaten. Die nötige Konfiguration von Datentypen, das Einbinden des Decoders in die "mainFunktion sowie die Anpassung des Scatterplots und der Punkte (richtige Farbe, Anzeige, Hover-Interaktion mit den Punkten) stellten eine Herausforderung, die durch "Try and Error"bewältigt werden konnte. Die gemachten Erfahrungen mit der Verknüpfung zwischen Decoder und Codegrundgerüst konnten gut auf die Programmierung der Parallelen Koordinaten übertragen werden, sodass die Grundfunktionsweise (keine Interaktionsmöglichkeit mit der Anwendung) erfolgreich übertragen werden konnte. Nach Tests zu der Röntgendarstellung der Linien im Vergleich zu "normaler" Darstellung wurde die Röntgendarstellung für die zweite Visualisierung entschieden. Für die Baumhierarchie stellte die erste Phase die Datenvorverarbeitungsschritte für die JSON-Datei dar. In der Übung 10 wurde ebenfalls ein JSON-Decoder verwendet, sodass keine nennenswerten Komplikationen mit dem Lesen der JSON-Datei auftraten. In der zweiten Phase wurde das default-Layout durch einen eigene Layout-Funktion ersetzt, wozu zunächst der Elm-Paketimport erweitert wurde. Die dritte Phase umfasste das Programmieren von Interaktionsmöglichkeiten mit der Anwendung, wie die Buttons zum Wechseln der Y-Achsen beim Scatterplot und den Achsentauch der Parallelen Koordinaten. Eine Herausforderung stellte die dafür notwendige Anpassung der Programmgliederung dar, was nicht nur durch die oben genannten Übungen als Anhaltspunkt, sondern ebenfalls durch Ausprobieren und Recherche in den Elm-Dokumentationen funktionierte. Um die drei Visualisierungsdateien miteinander zu verbinden, wurde der `elm-make` Befehl verwendet und die Dateien untereinander und mit einer `index.html` Seite verknüpft, die als Startseite der Anwendung dient. Hierbei war eine regelmäßige Iteration notwendig, um in den jeweiligen Programmen vorgenommene Änderungen zu aktualisieren.

5 Anwendungsfälle

Im folgenden werden für die drei Darstellungen in Hinblick auf die Zielgruppe praktische Anwendungsfälle aufgezeigt.

5.1 Anwendung Visualisierung Eins

In der Abbildung 5 ist ein Scatterplot mit der Achsenauswahl Alter zu Autos abgebildet, welcher sich besonders für die Zielgruppe des Fahrradhandels richtet.

Aus dieser Konfiguration lassen sich einige Informationen entnehmen, welche dabei helfen können, die Kundenzielgruppe besser einzuordnen. Aus diesen Informationen können beispielsweise im Rahmen von Schulungen Fragen abgeleitet werden, deren Antworten dabei helfen können, Kaufinteressierte besser einzuordnen. So lässt sich erkennen, dass hauptsächlich Personen zwischen Ende 20 und Mitte 50 kein Auto und mehrheitlich ein Fahrrad besitzen. Diese Altersgruppe entspricht demographisch auch stark der berufstätigen Bevölkerung. Des weiteren haben Personen mit vier Autos tendenziell weniger Fahrräder. Auffällig wenige befragte Personen mit zwei Autos haben kein Fahrrad. Bei einem und bei drei angegebene Autos ist die Anzahl ohne Fahrrad



Abbildung 5: Anwendung Scatterplot, Quelle: Eigene Darstellung

zwar höher, allerdings ist der Vergleich ausgewogener als bei drei Autos. Personen mit vielen Autos werden voraussichtlich entsprechende Stellplätze zur Verfügung haben, wo sie platzsparender für verschiedene Einsatzbereiche (wie Baustelle im Stadtverkehr, bei einem Auto und Autoreparatur) ein Fahrrad als Alternative zum Auto abstellen könnten. Diese Personengruppe mit vier Autos hat bis auf eine Ausnahme Berufe im Management und als spezialisierte Fachkräfte. Bei den Personen ohne Auto sind die Berufsgruppen gemischt, wobei viele Büroangestellte oder tätig im Handwerk sind. Durch die Zusatzinformation zum Eigenheimbesitz können ebenfalls Rückschlüsse auf Einkommen und die Bereitschaft mehr für ein Fahrrad auszugeben und potenziell mehr Platz für mehr Fahrräder mit verschiedenen Verwendungszwecken gezogen werden. Hier besteht für den Fahrradhandel die Möglichkeit Rabatte auf den Kauf weiterer Fahrräder für weitere Verwendungsbereiche anzubieten.

Um diese Zusammenhänge aufzuzeigen kommt als Alternative nur bedingt die Darstellung über Parallele Koordinaten in Frage. Über die stärker hervortretenden Polygonlinien der Röntgendarstellung ist erkennbar, dass jüngere Leute überwiegend kein Auto haben. Da hier allerdings keine Werte eindeutig getrennt voneinander betrachtet werden können. Darüber hinaus sind über die verschiedenen Farbakzente der Punkte und Zusatzinformationen beim Hovern Funktionsweisen mit dem Scatterplot verbunden, die über keine Darstellungsalternative zu ähnlich qualitativen Ergebnissen kommen würden.

5.2 Anwendung Visualisierung Zwei

Die zweite Anwendung stellt die in Abbildung 6 gezeigte Achsenumstellung dar. Diese ist besonders für Fahrradhersteller geeignet, weist aber auch Eigenschaften auf, die für den Fahrradhan-

del interessant sind. Im dazugehörigen Elm-Programmcode wurden Daten herausgefiltert, die "neinßum Fahrradkauf und als Region nicht Europa angeben haben. Durch die erste Filterung lässt sich für die Fahrradhersteller besser erkennen, was nachgefragt wird. Die zweite Vorfilterung wurde vorgenommen, damit der Datensatz auf Grund zu vieler Daten nicht zu unübersichtlich wird. Die meisten Datenwerte sind für die Region North America verfügbar. Danach folgen in Abstufender Reihenfolge Europe und "Pacific". Europa wurde auf Grund des Regionalbezugs und der damit einhergehenden Assoziation für Regionale Bauunternehmen gewählt.

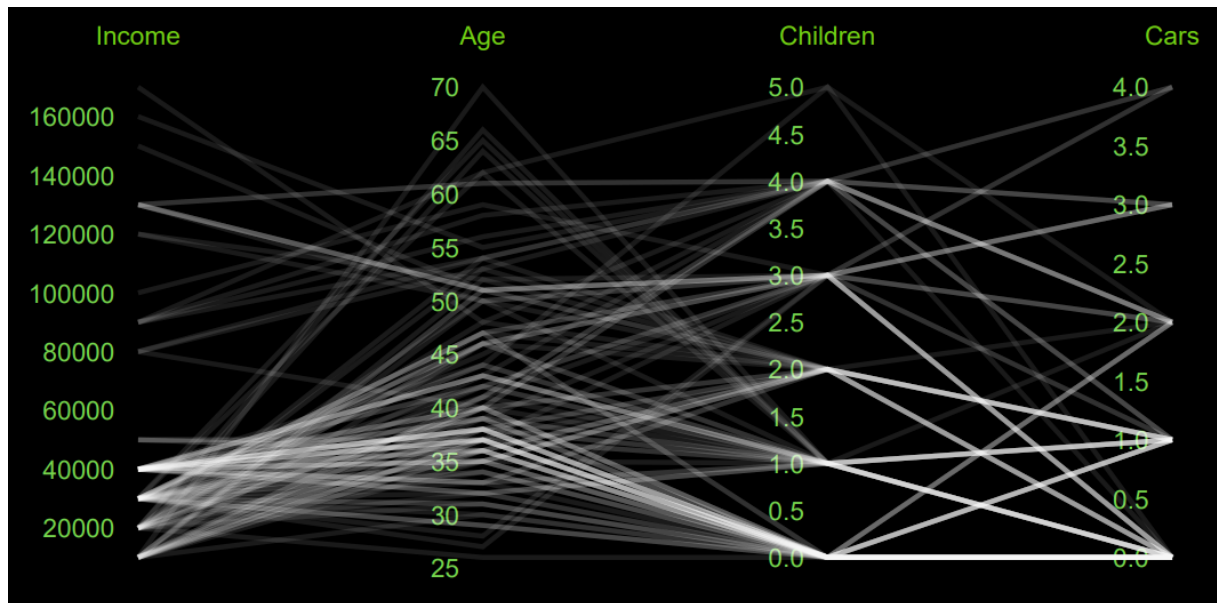


Abbildung 6: Anwendung Parallele Koordinaten, Quelle: Eigene Darstellung

In dieser Abbildung lässt sich erkennen, dass die Mehrheit, der vorgefilterten Fahrraddaten ein im Vergleich mit der Skala von 10.000 bis 170.000 geringes Einkommen aufweisen und dass dies mit zunehmendem Alter steigt. Überwiegend jüngere Personen zwischen 30-40 haben keine Kinder. Ältere Personen haben tendenziell mehr als ein Kind und würden dementsprechend auch Kinderfahrräder kaufen, was für Fahrradhändler eine Information mit Mehrwert bedeutet. In Hinblick auf die in 1.1 geschilderten Zusammenhänge können Fahrradhersteller für ältere Personen höherpreisige Fahrräder anbieten, welche die angesprochenen Eigenschaften, wie bessere Abstiegsmöglichkeiten durch niedrigere Gabel oder schmalere Griffe besitzen und neue Technologien verwenden. Im Umkehrschluss bedeutet dies, dass Hersteller für eine jüngere Zielgruppe preiswertere Fahrräder herstellen sollten, da diese, wie in Abbildung 6 ersichtlich mehr jüngere Fahrradkaufende als ältere beinhaltet.

Durch die Parallelen Koordinaten können übersichtlich Trends und Verallgemeinerungen abgeleitet werden, bieten aber gleichzeitig die Möglichkeit unwichtige Bereiche anzuschauen, weshalb diese Darstellung für die multidimensionalen Zahlenwerte der Fahrraddaten am besten funktioniert. Diese intuitive Beobachtung ist am ausdrucksstärksten in Kombination mit dem

Scatterplot. Durch den Direktvergleich der beiden Visualisierungen fusionieren die Stärken beider Anwendungen. Beispielsweise lassen sich Muster erkennen, wie dass Personen mit 4 bis 5 Kindern ein höheres Einkommen haben. Für Fahrradhersteller ist außerdem die genaue Hoveranzeige in Verbindung mit dem Scatterplot interessant, da hier zusätzlich Informationen zum Geschlecht eingeholt werden können.

5.3 Anwendung Visualisierung Drei

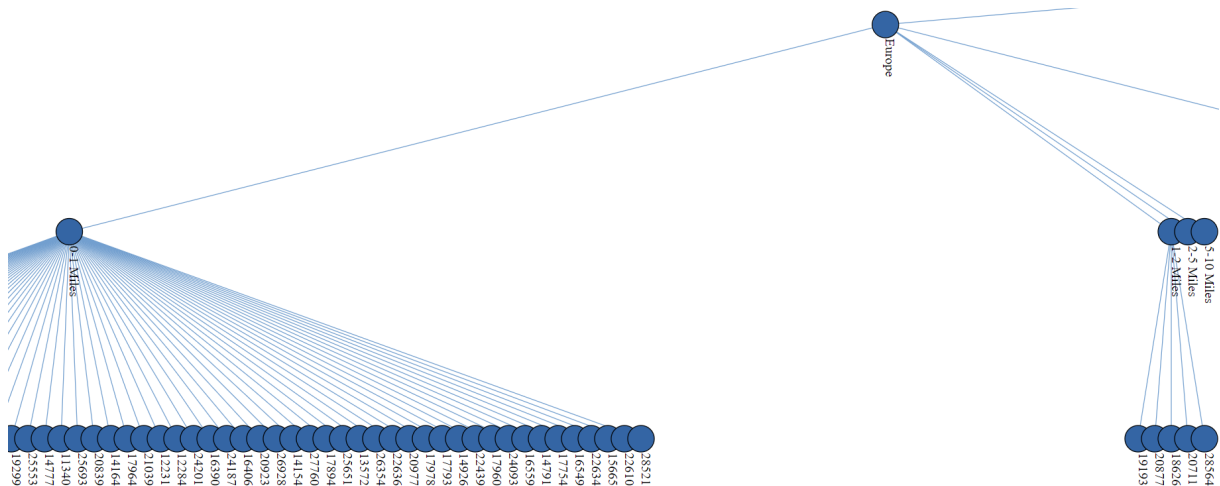


Abbildung 7: Anwendung Baumhierarchie, Quelle: Eigene Darstellung

Aus der Abbildung 7 geht hervor, dass in Europa von den Fahrradkäufern ohne Auto die überwiegende Mehrheit eine geringe Distanz von Zuhause zur Arbeit aufweist. Da diese Distanz für den öffentlichen Nahverkehr zu gering wäre, gleichzeitig aber Distanzen von 1.6 km nicht mehr zeitsparend und ohne Anstrengung zu Fuß zu bewältigen sind, kann davon ausgegangen werden, dass viele dieser Fahrradkaufenden das Fahrrad zur Arbeit verwenden. Auf Grund der kurzen Entfernung ist davon auszugehen, dass die Daten in Großstädten gesammelt wurden. Die Infrastruktur ist hiervon häufig von mehrspurigen Straßen mit vielen unübersichtlichen Kreuzungen und Abbiegungen gekennzeichnet. Infrastrukturprogramme, wie in der Einleitung beschrieben, sehen hierzu vor, dass besonders Fahrradwege besser ausgebaut werden sollen, damit weniger Verkehrsunfälle mit Fahrrädern passieren. Gerade deshalb ist dieser Ausschnitt aus der Baumhierarchie für in Frage kommende Bauunternehmen spannend. Daran lässt sich erkennen, dass vermeintlich kurze Distanzen zwischen bereits vorhandener Infrastruktur gebaut werden müssen, um zu einer klimaneutralen, sicheren und verkehrsberuhigten Innenstadt beizutragen. Andere Darstellungsmethoden, die für eine vergleichbare Darstellung in Frage kämen, wie die Graphendarstellung, würden zur Unübersichtlichkeit der Verzweigungen auf Grund der großen Datenmenge beitragen. Deshalb ist die Baumhierarchiestruktur für diesen Sachverhalt nämlich das Aufzeigen der verschiedenen Pendeldistanzen optimal gewählt.

6 Verwandte Arbeiten

Zum Themenbereich Fahrrad und damit verbundenen Entwicklungen wird eine Auswahl an Publikationen vorgestellt, welche vergleichbare Visualisierungstechniken zu diesem Forschungsbericht verwendet haben.

Eine Forschungsarbeit zum Thema Nutzungsverhalten in Verbindung mit E-Bikes greift unter anderem auf Scatterplots zurück, um Zusammenhänge zwischen mit E-Bikes zurückgelegten Kilometern und gutem Wetter aufzuzeigen [29]. Ein weiteres Projekt zur Untersuchung des Fahrradverleihs im südkoreanischen Seoul zeigt unter Einsatz der Scatterplot-Technik, dass ein Zusammenhang zwischen hohen Temperaturen und erhöhter Anzahl an vermieteten Fahrrädern besteht und dass bei hohen Windgeschwindigkeiten weniger Fahrräder ausgeliehen werden [30]. Gemeinsamkeiten zu diesen beiden Arbeiten liegen in der Verwendung der gleichen Visualisierungstechnik. Anders als bei den verwandten Arbeiten wird in dieser Arbeit auch die Interaktion mit der Visualisierung ermöglicht. Des weiteren existiert eine Arbeit zur Informationsvisualisierung, die sich auf die Visualisierung von Gemeinsamkeiten und Unterschieden zwischen Fahrradverleih und Taxiservice konzentriert. Hierbei wird eine Anwendung bereitgestellt, mit der interagiert werden kann. Über Buttons kann zwischen einem Hierarchieüberblick, Liniengraphen zu Zeitverläufen und zwei Stadtkarten-Heatmaps für Serviceregionen die Ansicht gewechselt werden. Für den Hierarchieüberblick wurde anstelle eines Baumdiagramms ein Sunburst-Diagramm verwendet, da hierbei ein sehr großer Datensatz mit wenig Platz abgebildet werden kann. Neben den erwähnten Buttons ist die Interaktion mit der Vergleichsanwendung tiefgreifender. So haben Anwendende die Möglichkeit, in der Hierarchie zu filtern, Muster direkt auszuwählen, und die Daten über eine Navigationszeile in eine beliebige Reihenfolge bringen. Darüber hinaus können erkannte Muster erstellt, benannt und andere Muster gelöscht werden. Über Mausinteraktionen kann aus der Anwendung heraus und hineingezoomt werden. Ähnlich wie die Hoverfunktion bei den Scatterplots dieser Arbeit kann beim Hovern über die Sunburst Hierarchie der Sektor hervorgehoben werden. Diese Forschungsarbeit weist detailliertere Interaktionsmöglichkeiten als diese Forschungsarbeit zu den Fahrradkaufdaten auf. Ein Ziel dieser verwandten Arbeit von Dai et al. [31], nämlich die Stadtplanung zu verbessern, überschneidet sich mit dem Visualisierungsziel des Baumdiagramms dieser Arbeit, die Planung der Fahrradweginfrastruktur von involvierten Bauunternehmen zu unterstützen.

Vom Themengebiet zu Fahrrädern im engen Sinne entfernt, zum Thema Wettkampfsport visualisieren mit Hilfe von Parallelen Koordinaten die Daten zur Rugby-Sportart [32, 33]. Eine weitere Studie zu öffentlichen Fahrradverleih-Systemen, mit Interaktionsmöglichkeiten verwendet verschiedene Weiterentwicklungsmöglichkeiten der Parallelen Koordinaten. In der Standardvariante werden ebenfalls Attribute gegenübergestellt. In der beschriebenen Arbeit werden allerdings fünf Achsen verwendet. Mit deren Anwendung können Nutzer eigenständig Daten filtern und durch Hovern über die Linien welche die Parallelen Koordinaten verbinden, werden die jeweiligen Linien hervorgehoben. Grundsätzlich gemeinsam mit diesem Ansatz, ist dass die Daten gefiltert

werden. Der Unterschied in dieser Arbeit ist, dass die Daten im Elm-Code vorgefiltert wurden und nicht durch Visualisierungsanwendende. Somit stellen die Unterschiede die Filteranwendung durch Nutzende und das Hovern dar [34].

7 Zusammenfassung und Ausblick

Dieses Projekt konnte die Datenaufbereitung zu einem sehr aktuellen Thema visuell umsetzen. Diese Aktualität wird durch die vermehrte Ausstellung von Fahrrädern auf der zeitgleich zur Projektabgabe stattfindenden IAA Mobility in München verdeutlicht [35].

Durch den Scatterplot konnten Zusammenhänge aufgezeigt werden, welche insbesondere für die Zielgruppe des Fahrradhandels interessant sind. Trotz der Limitation, dass lediglich zwei Daten über die X und Y Achse gegenübergestellt werden, ermöglicht die Hover-Interaktion genügend Möglichkeiten für die Anwendenden, eigene Muster zu erkennen und für engere Kundenbindungen im Fahrradverkauf zu übertragen. Des Weiteren können Fahrradhersteller mit der Visualisierungstechnik der Parallelen Koordinaten interagieren sowie Informationen und Muster (z.B. Einkommen, Alter) erkennen, welche bei der Fahrradproduktion unterstützen können.

Die dritte Visualisierung ermöglicht es einen komplexen Sachverhalt übersichtlich und ohne notwendige Vorkenntnisse aufzuzeigen, der intuitiv verständlich ist. Über die Baumhierarchie wird die Relevanz für kurze Fahrradwege sichtbar. Besonders in Großstädten ist diese Thematik brisant, da hier im Verkehr häufig Fahrradunfälle passieren [36, 37, 38]. deshalb eignet sich diese Visualisierung für Bauunternehmen im Bereich der Fahrradwege-Infrastruktur und alle an der Planung und Bau beteiligten Gruppen.

Somit löst dieses Projekt die Herausforderung den Anforderungen mehrerer Zielgruppen gleichzeitig gerecht zu werden. Im Bezug auf Interaktion stellen folgende Inhalte eine Weiterentwicklung dar. Die Hauptwebsite `index.html` könnte noch ansprechender visualisiert werden. Einen weiteren Punkt stellen die Interaktionsmöglichkeiten für Anwendende dar, welche so erweitert werden könnten, dass Nutzende selbst die Möglichkeit haben die Daten nach Region, bestimmten Werten oder Wertschwellen, oder individuellen Parametern über Buttons zu filtern. Dies würde die Zielgruppe nochmals erweitern, sodass nicht nur europäische sondern auch amerikanische oder pazifische Branchenunternehmen angesprochen werden. Die Interaktion im Bereich der Parallelen Koordinaten könnte ebenfalls Hoverfunktionen beinhalten, wozu die `parallelCoordinatesPlot`-Funktion im in Bereich angepasst werden müsste. Im Rahmen dieses Projekts wurde auf diese Anpassung verzichtet, da Zusatzinformationen über das Hovern im Scatterplot verfügbar sind.

Aus Visualisierungssicht könnten mehr Visualisierungstechniken wie die Sternenkoordinaten oder bei der Erhebung zeitlich abhängiger Fahrradaten (gefahrte Kilometer pro Woche, etc.) Zeitdiagramme verwendet werden, welche als zusätzliche Interaktionsmöglichkeit verknüpft sein sollten. Auch eine Option zum Ein- und Ausschalten der Röntgenansicht könnte für die Parallele Koordinaten sinnvoll sein. Auf Datenebene könnten Erhebungen zur Häufigkeit des Fahr-

radfahrens, verwendetem Equipment wie Helme oder zusätzliche Beleuchtung, Verkehrssicherheit, Fahrstil interessant sind, um zusätzliche Gadgets zu verkaufen oder auch die Zielgruppe der Versicherungsbranche anzusprechen. Diese Daten könnten durch Umfragen erhoben und die Visualisierungen im Rahmen zukünftiger Forschungsprojekte erweitert werden.

Zusammenfassend bereitet diese Forschungsarbeit zur Visualisierung von Fahrradkäufen im B2C-Segment für die Anbieterseite über die verschiedenen Darstellungstechniken Zusammenhänge, wie Einkommen, Alter und die Autoanzahl, die zum Fahrradkauf führen auf. Wie durch die Baumdarstellung ersichtlich, stehen mit dem Fahrradboom besonders Kurzstrecken im Fokus, welche im Zusammenhang mit Infrastrukturprogrammen ausgebaut werden sollten. Mit der verfügbaren Zeit für die Projektarbeit, der zur Verfügung stehenden Ressourcen und Knowhow konnten für die Zielgruppen detailliertere Informationen visualisiert werden, wie Pendeldistanzen und die Betrachtung von Nichtfahrradbesitzenden als potenzielle Neukunden, die über einen groben Überblick hinausgehen.

Anhang: Git-Historie

Literatur

- [1] Heike Marquart, Julia Schuppan, Benjamin Heldt, Lisa Buchmann, Julia Jarass, Sarah Berg, Till Steinmeier, Philipp Masius, Meret Nathalie Batke, Arthur Zschäbitz, Jakob Bastian, Charlotte Blechner, David Brunner, Julian Maurer, Pascal Kraft, Leon Govinda Stephan, Tuan Anh Rieck, Konstantin Arndt, Lennart Goettsche, Robert Radloff, Nadja Martin, Lara Ann Steinert, and Fabian Drews. *Mobilität in Stadtquartieren*. Humboldt-Universität zu Berlin, 2021. doi: 10.18452/22930.
- [2] Martin Kords. Statistiken zum thema fahrradfahrer, 2020. URL <https://de.statista.com/themen/6616/fahrradfahrer/>.
- [3] Statista. Corona-krise sorgt für fahrrad-boom, 25.08.2021. URL <https://de.statista.com/news/fahrradfahrer/de>.
- [4] Martin Platter. Das virus bewegt aufs velo, 2020. URL <https://www.nzz.ch/mobilitaet/fahrraeder-sind-dank-corona-beliebt-wie-selten-zuvor-ld.1560377>.
- [5] Christiane Köllner. Individualverkehr gewinnt, öpnv verliert, 2020. URL <https://www.springerprofessional.de/mobilitaetskonzepte/verkehrswende/individualverkehr-gewinnt--oepnv-verliert/18656208>.
- [6] Christian Thomann-busse. Setzt sich der fahrradtrend weiter so fort?, 2020. URL <https://www.zdf.de/nachrichten/wirtschaft/fahrradbranche-boom-jahr-2020-100.html>.
- [7] Sinus Markt- und Sozialforschung GmbH and Bundesministerium für Verkehr und digitale Infrastruktur. *Fahrrad-Monitor Deutschland Corona-Befragung 2020: Ergebnisse einer repräsentativen Online-Befragung*. SINUS-Institut, Heidelberg, 2021. URL https://www.bmvi.de/SharedDocs/DE/Anlage/StV/fahrrad-monitor-deutschland-corona-befragung-2020.pdf?__blob=publicationFile.
- [8] ADAC.de. Neu- und ausbau: Millionen für deutsche radwege | adac, 2021. URL <https://www.adac.de/news/radwege-ausbau/>.
- [9] Muenchen.de. Neue fahrradwege in münchen: Arbeiten abgeschlossen, 2021. URL <https://www.muenchen.de/verkehr/aktuell/pop-up-rad-wege-muenchen.html>.
- [10] Andreas Jöhrens. Boomendes geschäft, steigende preise: Lieferprobleme im fahrrad-handel, 2021. URL <https://www.tagesschau.de/wirtschaft/verbraucher/fahrrad-e-bike-boom-101.html>.
- [11] N-tv Nachrichten. Sportliche premium-fahrräder: Cyklaer - porsche baut jetzt pedelecs. *n-tv NACHRICHTEN*, 31.08.2021. URL <https://www.n-tv.de/auto/Cyklaer-Porsche-baut-jetzt-Pedelecs-article22769128.html>.

- [12] Forschungsinformationssystem. Käuferprofil von elektrofahrradnutzern, 2017. URL <https://www.forschungsinformationssystem.de/servlet/is/355199/>.
- [13] Fahrrad XXL Blog. Fahrrad xxl e-bike-studie 2020 - fahrrad xxl blog, 2020. URL <https://www.fahrrad-xxl.de/blog/e-bike-studie/#5>.
- [14] Radfahren.de. Damenfahrrad: Welche vorteile bietet ein spezielles fahrrad für frauen?, 2018. URL <https://www.radfahren.de/service/damenfahrrad-vorteile/>.
- [15] Tretwerk. Welches fahrrad für welchen zweck? – tretwerk biker blog, 2018. URL <https://tretwerk.net/blog/welches-fahrrad-fuer-welchen-zweck/>.
- [16] Mike Yi. A complete guide to scatter plots, 2019. URL <https://chartio.com/learn/charts/what-is-a-scatter-plot/>.
- [17] F. J. Anscombe. Graphs in statistical analysis. *The American Statistician*, 27(1):17, 1973. ISSN 00031305. doi: 10.2307/2682899.
- [18] William S. Cleveland and Robert McGill. The many faces of a scatterplot. *Journal of the American Statistical Association*, 79(388):807, 1984. doi: 10.2307/2288711.
- [19] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In IEEE, editor, *Proceedings of the First IEEE Conference on Visualization: Visualization '90*, pages 361–378. IEEE Comput. Soc. Press, 1990. ISBN 0-8186-2083-8. doi: 10.1109/VISUAL.1990.146402.
- [20] Rida Moustafa and Ed Wegman. Multivariate continuous data — parallel coordinates. In *Graphics of Large Datasets*, Statistics and Computing, pages 143–155. Springer New York, New York, NY, 2006. ISBN 978-0-387-32906-2. doi: 10.1007/0-387-37977-0{\textunderscore}7.
- [21] Julian Heinrich and Daniel Weiskopf. Parallel coordinates for multidimensional data visualization: Basic concepts. *Computing in Science & Engineering*, 17(3):70–76, 2015. ISSN 1521-9615. doi: 10.1109/MCSE.2015.55.
- [22] Stephen Few. Line graphs and irregular intervals. *Visual Business Intelligence Newsletter*, (11):1–11, 2008. URL http://www.perceptualedge.com/articles/visual_business_intelligence/line_graphs_and_irregular_intervals.pdf.
- [23] Julian Heinrich and Daniel Weiskopf. Continuous parallel coordinates. *IEEE transactions on visualization and computer graphics*, 15(6):1531–1538, 2009. ISSN 1077-2626. doi: 10.1109/TVCG.2009.131.
- [24] Heinz-Peter Gumm and Manfred Sommer. *Programmierung, Algorithmen und Datenstrukturen*, volume / Heinz-Peter Gumm, Manfred Sommer ; Band 1 of *De Gruyter Studium*.

- De Gruyter Oldenbourg, Berlin and Boston, 2016. ISBN 978-3-11-044227-4. URL http://www.degruyter.com/search?f_0=isbnissn&q_0=9783110442274&searchTitles=true.
- [25] VeloTOTAL - Das größte Netzwerk rund um das Thema Fahrrad. Click & meet, click & collect, click & ride safe – so katapultieren sich berliner fahrradlädern ins digitale zeitalter, 2021. URL <https://www.velototal.de/2021/03/09/click-meet-click-collect-click-ride-safe-so-katapultieren-sich-berliner-fahrradl%C3%A4den-ins-digitale-zeitalter/>.
- [26] Sebastian Follmer. Fahrrad online-shops: Bike-versender im vergleich, 2015. URL https://praxistipps.chip.de/fahrrad-online-shops-bike-versender-im-vergleich_9611.
- [27] Kevin Knitterscheidt and Silke Kersting. Rohstoffmangel: Materialknappheit bedroht bau-boom, 2021. URL <https://www.handelsblatt.com/unternehmen/industrie/holz-stahl-kupfer-der-rohstoffmangel-bedroht-den-bau-boom/27284062.html?ticket=ST-22359-oS7wwMQLMxJJ3LFaIOY0-ap2>.
- [28] Heeral Dedhia. Bike buyers 1000, 22.09.2020. URL https://www.kaggle.com/heeraldedhia/bike-buyers?select=bike_buyers.csv.
- [29] Ivan Rios, Lukasz Golab, and S. Keshav. Analyzing the usage patterns of electric bicycles. In *Proceedings of the Workshop on Electric Vehicle Systems, Data, and Applications*, pages 1–6, New York, NY, USA, 06212016. ACM. ISBN 9781450344203. doi: 10.1145/2939953.2939955.
- [30] Aditya Singh Kashyap and Swastika Swastik. Regression model to predict bike sharing demand. *International Journal of Innovative Science and Research Technology*, 6(3):1024–1028, 2021.
- [31] Haoran Dai, Yubo Tao, and Hai Lin. Visual analytics of urban transportation from a bike-sharing and taxi perspective. *Journal of Visualization*, 23(6):1053–1070, 2020. ISSN 1343-8875. doi: 10.1007/s12650-020-00673-8.
- [32] Meng Du and Xiaoru Yuan. A survey of competitive sports data visualization and visual analysis. *Journal of Visualization*, 24(1):47–67, 2021. ISSN 1343-8875. doi: 10.1007/s12650-020-00687-2.
- [33] David H. S. Chung, Matthew L. Parry, Iwan W. Griffiths, Robert S. Laramee, Rhodri Bown, Philip A. Legg, and Min Chen. Knowledge-assisted ranking: A visual analytic application for sports event data. *IEEE computer graphics and applications*, 36(3):72–82, 2016. doi: 10.1109/MCG.2015.25.
- [34] Xiaoying Shi, Zhenhai Yu, Jing Chen, Haitao Xu, and Fei Lin. The visual analysis of flow pattern for public bicycle system. *Journal of Visual Languages & Computing*, 45:51–60, 2018. ISSN 1045926X. doi: 10.1016/j.jvlc.2017.03.007.

- [35] IAA.de. Radfahren als elementarer bestandteil des mobilitätskonzepts der iaa, 08.09.2021. URL <https://www.iaa.de/de/mobility/aussteller/aussteller-werden/vorteile-formate/fahrrad-bei-der-iaa-mobility-2021>.
- [36] Felix Reek. Adfc-fahrradklima-test: In karlsruhe fühlen sich radler besonders wohl, 2021. URL <https://www.sueddeutsche.de/auto/fahrrad-verkehr-umfrage-adfc-1.5236313>.
- [37] Joanna Nogly. Adac testet radverkehr in großstädten: 354 tödliche unfälle, 2014. URL <https://taz.de/ADAC-testet-Radverkehr-in-Grossstaedten/!5034472/>.
- [38] tagesschau. Unfallstatistik: Zahl der verkehrstoten auf 30-jahres-tief, 2021. URL <https://www.tagesschau.de/inland/verkehrstote-deutschland-103.html>.