

BSc Data Science

GDV-Report 2022

Untersuchungen über die Sozialhilfeempfänger-Quoten der Stadt Basel

Yannic Lais
26.5.2022

Inhalt

Einleitung.....	2
LE1: Visualization basics, chart types	2
LE 2: Visual Perception	4
LE 3: Design Principles vs. Data	6
LE 4: Grammar of Graphics Tools	8
LE 5 Evaluation:	10
Anhang	12
Datensatz.....	12
Quellen	12
Allgemein.....	12
LE1	12
LE2	12
LE3	12
LE4	12
LE5	12

Einleitung

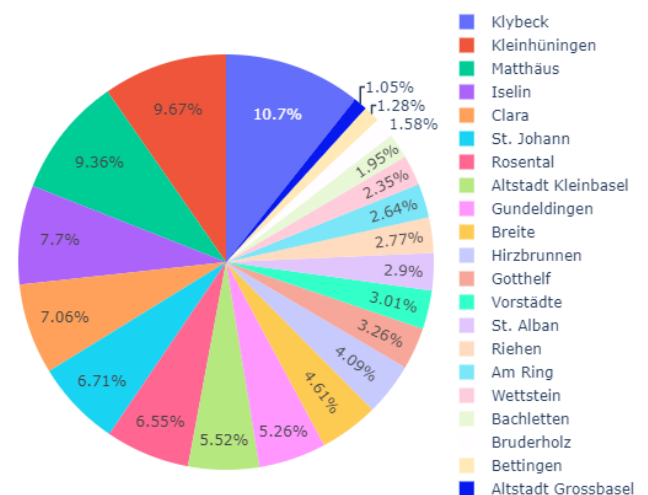
Die Sozialhilfe in der Schweiz unterstützt Menschen in Notlagen, welche die Sozialversicherungen nicht abdecken. Beispielsweise wenn nach einer Scheidung zu wenig Geld vorhanden ist oder aber auch nach längerer Zeit ohne Arbeit. Die Sozialhilfe ist die Stütze unserer Gesellschaft und ein Auffangbecken, aus welchem man sich baldmöglichst wieder hinauswinden sollte. In folgendem Bericht erfahren Sie mehr über Korrelationen zwischen einem hohen Anteil an Sozialhilfeempfänger und den Eigenschaften der Stadtviertel von Basel. Wenn in diesem Bericht von Sozialhilfe gesprochen wird, ist dabei die Sozialhilfe im engeren Sinn gemeint. Von der Sozialhilfe im engeren Sinn ist die Rede, wenn es sich um aktive Armutsbekämpfung einzelner Personen handelt, die Leistungen also über Stipendien, Kindergeld, AHV/IV Ergänzungsleistungen herausgehen [1,2]. Die Daten wurden von dem Kanton Basel-Stadt zur Verfügung gestellt. Die Untersuchungen und Visualisierungen werden hauptsächlich in Python mit den Libarys Pandas, Matplotlib, Plotnine und Plotly erstellt.

LE1: Visualization basics, chart types

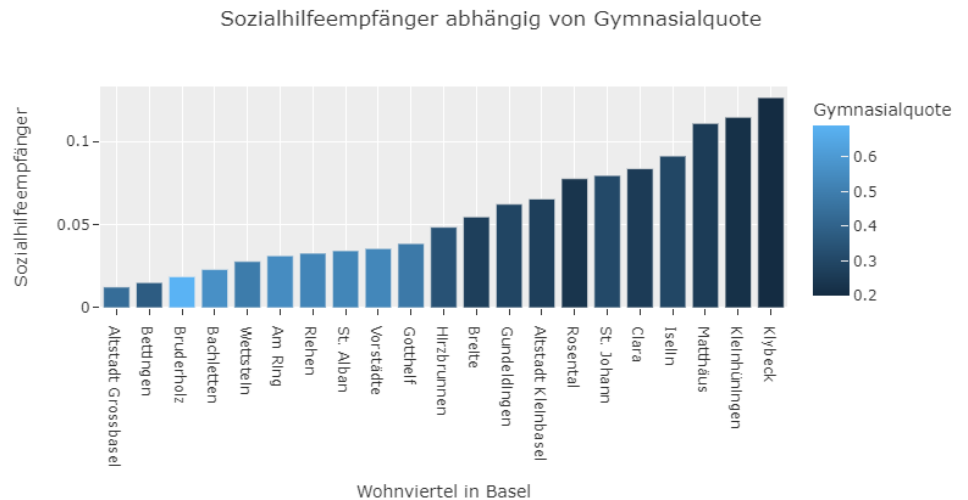
Um Korrelationen und Veränderungen in Daten einfach und verständlich darzustellen sind Visualisierungen perfekt geeignet. Dabei bietet sich eine breite Palette an verschiedenen Charts an [5,6]. Wichtig dabei ist, seine Visualisierungen auf das Publikum anzupassen. So sollte man bei Personen mit wenig Kenntnissen einfache Charts wie das Histogramm oder einen Pie-Chart verwenden. Das Pie-Chart eignet sich hervorragend, um Proportionen darzustellen. Um Veränderungen über einen gewissen Zeitraum darzustellen, eignen sich Histogramme oder Linien Charts [7,8]. Um Korrelationen oder Cluster hervorzuheben sind Scatterplots gut geeignet. Im nächsten Abschnitt sehen Sie einige Beispiele aus dem Datensatz.

Um einen Überblick zu erhalten, wurden die Anteile an Sozialhilfebezügern in den Quartieren mittels eines Pie-Charts visualisiert. Zu sehen sind grosse Unterschiede zwischen den einzelnen Quartieren. So sind in den Quartieren Klybeck und Kleinhüningen im Jahr 2021 rund 10% der Einwohner von der Sozialhilfe abhängig, während in den Quartieren Altstadt Grossbasel, Bettingen, Bruderholz und Bachletten unter 2% von der Sozialhilfe unterstützt, werden müssen. Der Schweizerische Durchschnitt liegt bei 3,2%, wobei der Schnitt mit steigender Einwohnerzahl zunimmt. Dies ist sicherlich ein schlechtes Beispiel für die Verwendung eines Pie-Charts. Denn es werden zu viele und vor allem ähnliche Abschnitte miteinander verglichen, so ist der Unterschied von den Quartieren Clara (7.06%) und Altstadt Kleinbasel (5.52%) kaum ersichtlich. Geeigneter wäre hier ein Histogramm, ähnlich wie dem folgenden ohne zusätzliche Information wie der Gymnasialquote.

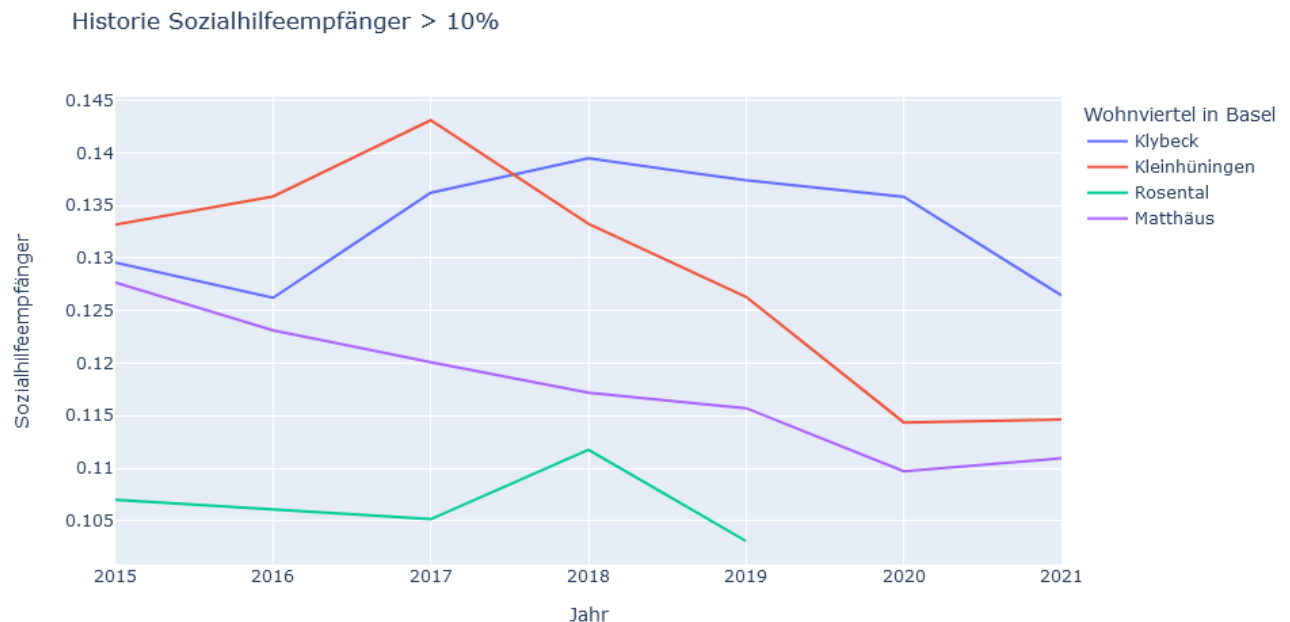
Verteilung Sozialhilfeempfänger 2021



Ein Grund für den Bezug von Sozialhilfe kann schlechte Chancen auf dem Arbeitsmarkt sein, welche oft durch fehlende Fort- und Weiterbildungen verstärkt werden. Im untenstehenden Histogramm ist deutlich zu sehen, wie die abnehmende Gymnasialquote mit dem zunehmenden Sozialhilfeempfänger in den einzelnen Quartieren korrelieren.



Ausserdem wurde untersucht, wie sich die Zahlen in Quartieren, mit besonders hohen Sozialhilfeempfängeranteilen, in den letzten Jahren veränderten. Spannend ist vor allem der Zeitraum 2019 bis und mit dem Jahre 2021 da wegen der Coronapandemie mit mehr Sozialhilfeempfänger gerechnet wurde. Diese These kann im Kanton Basel-Stadt jedoch nicht angenommen werden. Das Quartier 'Rosental' ist in den Jahren 2020 und 2021 nicht mehr zu sehen, weil es einen Sozialhilfeempfängeranteil unter 10% erreichte.

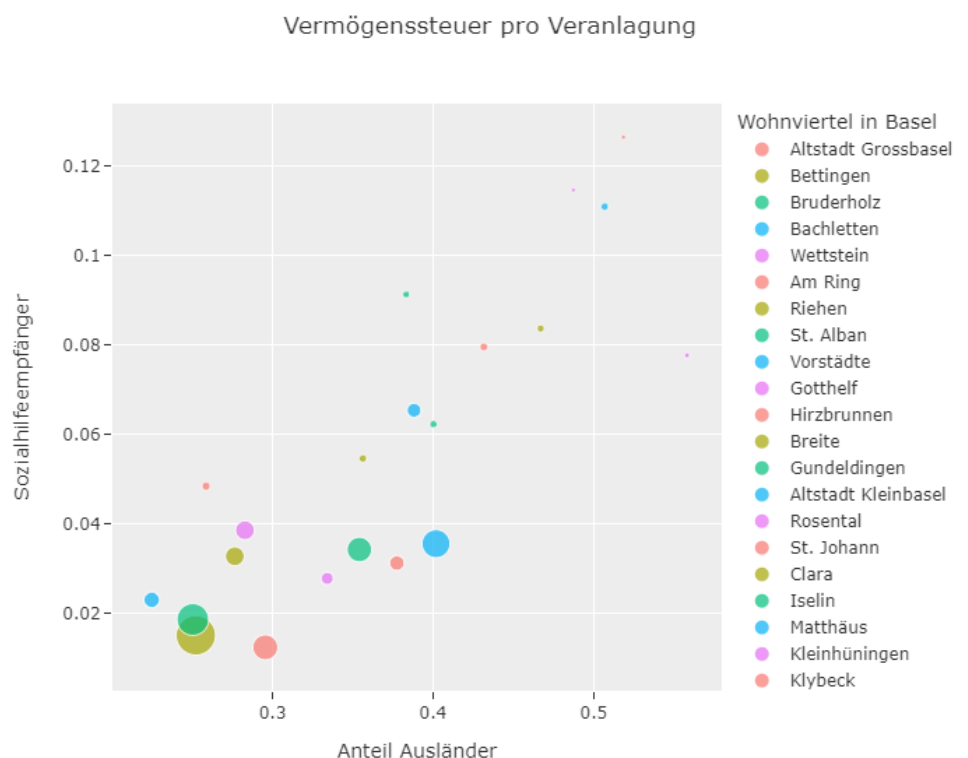


LE 2: Visual Perception

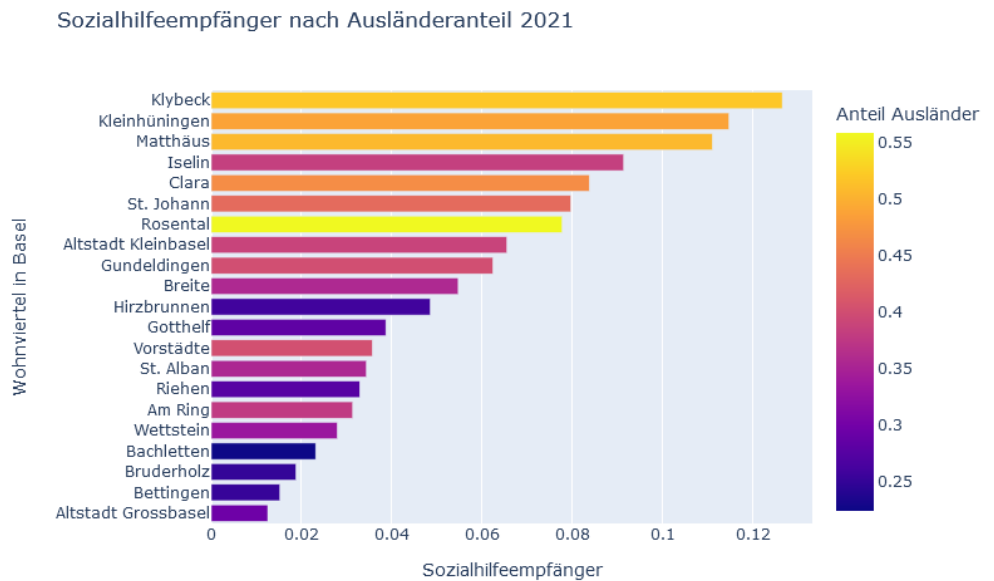
Daten gut und verständlich zu visualisieren ist eine Kunst. Es gibt viel zu beachten, wenn Daten verständlich übermittelt werden sollen. So ist es wichtig, sich jederzeit in die Lage der Zuschauer zu versetzen und sich selbst zu fragen, ob das gezeigte auch sinnvoll vermittelt wird. Ein guter Trick dabei ist, sich einige Versionen der Plots nur kurz anzuschauen und so erkennen ob in kurzer Zeit die Informationen klar ersichtlich sind. So sind schöne 3D-Plots doch meist eher ein Hindernis für die Kommunikation zwischen den Daten und dem Publikum [9]. Um Bestimmte Korrelationen oder Veränderungen hervorzuheben, sind Farben oder auch die Sichtbarkeit gute Mittel [10]. Hierbei gilt es zu unterscheiden, was wir genau hervorheben möchten, beispielsweise nutzen wir sehr unterschiedliche Farben, um Kategorien zu trennen. Gewisse Farben assoziieren wir auch mit bestimmten Dingen (Rot=negativ und Grün = positiv). Zudem können Zusätzliche Informationen mithilfe von Grössenunterschieden oder mittels Geografischen Karten in einem Plot visualisiert werden. In den untenstehenden Beispielen sehen wir verschiedene Möglichkeiten zusätzlich Daten einzubringen oder bestimmte Cluster sichtbar zu machen [11].

Zu den Risikogruppen gehören nebst den unter 18-jährigen und Einwohner von Gemeinden mit mehr als 100'000 Einwohner auch Ausländer. Im unteren Scatterplot sehen wir, wie ein erhöhter Ausländeranteil mit einem erhöhten Anteil an Sozialhilfeempfänger einher gehen. Des Weiteren sehen wir, dass in Quartieren von Einwohner mit hohem Vermögen auch weniger

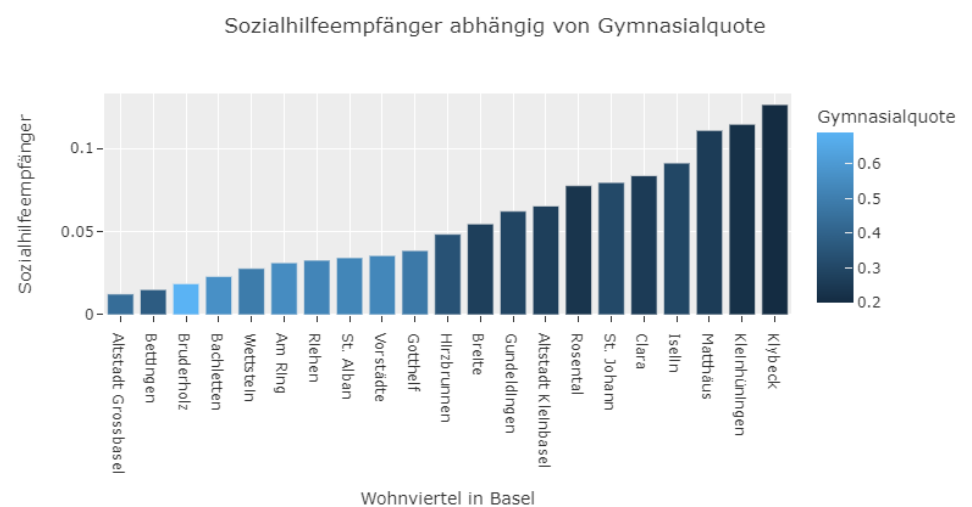
Sozialhilfeempfänger leben (Grösse der Punkte). So ist die durchschnittliche Vermögenssteuer pro Veranlagung in Bettingen rund 15'000 CHF und in Quartieren mit vergleichsweise hohem Ausländeranteil wie Klybeck bei 200 CHF. (Legende der Grösse momentan mit ggplot nicht möglich)



Wir wissen, dass nicht Schweizer aus verschiedenen Gründen zu einer Risikogruppe gehören. In diesem Plot sehen wir, dass in Quartiere mit höherem Ausländeranteil auch mehrheitlich mehr Sozialhilfeempfänger wohnen. Jedoch sehen wir auch, dass in Quartiere wie dem 'Hirzbrunnen' und dem 'Rosental' ähnlich viele Sozialhilfeempfänger wohnen, der Ausländeranteil jedoch sehr unterschiedlich hoch ist. In diesem Plot wurden sehr grelle Farben verwendet, deswegen sehen wir die unterschiedlichen Abstufungen sehr gut, zudem wurden bewusst keine Farben wie rot oder grün verwendet, da diese Farben meist positiv oder negativ in unseren Köpfen verankert sind.



Ein weiterer Grund für den Bezug von Sozialhilfe können schlechte Chancen auf dem Arbeitsmarkt sein, welche oft durch fehlende Fort- und Weiterbildungen verstärkt werden. Im untenstehenden Histogramm ist deutlich zu sehen, wie die abnehmende Gymnasialquote mit dem zunehmenden Sozialhilfeempfänger in



den einzelnen Quartieren korrelieren. Hierfür wurden dezenter Farben verwendet, zudem wurde für eine niedrige Gymnasialquote eine etwas aggressivere dunkle Farbe verwendet, welche dezent auf etwas negatives hindeuten soll. Ein weiteres Beispiel hierfür wären die Farben Rot und Blau, diese wurden jedoch bewusst nicht gewählt, da eine niedrige Gymnasialquote nicht per se als schlecht angeschaut werden kann oder soll.

LE 3: Design Principles vs. Data

Bei der Visualisierung der Daten ist es wichtig, saubere und gute Daten als Basis zu verwenden. Es wäre fatal, Entscheidungen aufgrund Visualisierungen zu treffen, welche aus falschen Daten hervorgegangen sind. Um solche Fehler zu vermeiden, gilt es einige Regeln zu beachten. Die erste Zeile sollte immer aus Titeln bestehen, so entsteht keine Verwirrung, was die Daten aussagen [12]. Zudem sollten die Attribute in jede Zeile vorhanden sein. Auch auf die Formate muss man achten und diese eventuell umformatieren [13]. Ausserdem müssen die verwendeten Daten auch korrekt sein und der Wahrheit entsprechen. Schon die Aufbereitung der Daten führt zu bestimmten Entscheidungen in der Visualisierung [14].

Der Datensatz, mit welchem die Visualisierungen gemacht wurden, stammt von Swissopendata des Kantons Baselstadt. Daher kann dem Datensatz und den darin enthaltenen Daten mit hoher Wahrscheinlichkeit auch vertraut werden. Der Datensatz ist schon sehr gut strukturiert. Einige Titel wurden zum vereinfachten Verständnis geändert. In dem Datensatz sind viele unterschiedliche Demografische-Daten der Bevölkerung, sowie auch Merkmale der einzelnen Quartiere, wie das durchschnittliche Baujahr eines Gebäudes. Da ich noch nicht wusste, nach welchen Merkmalen der Datensatz untersucht werden sollte, wurde der Datensatz nicht von einzelnen Spalten getrennt. Es waren auch keine grösseren Unregelmässigkeiten, bis auf einige fehlende Daten in gewissen Jahren, diese fehlenden Daten waren jedoch konstant und wurden in den bestimmten Jahren nicht erhoben. So war das Data-Wrangling eine einfachere Sache. Untenstehend sehen Sie einen kleinen Ausschnitt unserer Daten.

	Jahr	Wohnviertel_id	Wohnviertel in Basel	Sozialhilfeempfänger	Anteil Ausländer	Anteil Personen ohne Religionszugehörigkeit	Anteil Personen in Einzenhaushalten	Gymnasialquote	Altersquotient
142	2021	1	Altstadt Grossbasel	0.012505	0.295567	NaN	0.308090	0.437500	0.281538
16	2021	30	Bettingen	0.015190	0.252321	NaN	0.124545	0.375000	0.440909
144	2021	7	Bruderholz	0.018736	0.250450	NaN	0.147925	0.690323	0.412363
99	2021	8	Bachletten	0.023106	0.224846	NaN	0.211436	0.564356	0.360282
15	2021	14	Wettstein	0.027888	0.334142	NaN	0.272121	0.489362	0.283238

Für die meisten Visualisierungen wurde mit den Daten aus dem Jahr 2021 gearbeitet. Da sich die Sozialempfänger Rate über die Jahre bei einigen Quartieren stark verändert (vor allem in den letzten Jahren mit dem Corona-Virus) hat würde der Durchschnitt kein aktuelles Bild widerspiegeln, kann jedoch für einzelne Untersuchungen gut verwendet werden. Dafür musste man einige fehlende Daten wie die Religionsangehörigkeit und die Arbeitslosenquote in Kauf nehmen. Aus dem Datensatz wurden bewusst keine Spalten gelöscht, aus dem Grund, weil momentan noch keine offensichtlichen irrelevanten Daten bestimmt werden können.

Um ein Geomap-Plot zu erstellen, musste zusätzlich ein GeoJSON einbezogen werden. Dieses GeoJSON ist nötig, um die Grenzen und Formen der einzelnen Quartiere abzubilden. Ein GeoJSON ist ein auf JSON basierender Standard für Informationen, welche eine Geometrische Komponente besitzen. Es erlaubt uns Punkte, Linien oder eben Grenzen von Stadtviertel abzubilden. Aus dem GeoJSON mussten die jeweiligen Koordinaten in einer id klassifiziert werden. Anschliessend wurden die beiden Tabellen auf diese id's zusammengefügt [15].

Als nächstes möchten wir weitere Gründe für einen hohen Sozialhilfeempfängeranteil untersuchen. Die durchschnittliche Wohnfläche pro Person könnte ein Indikator für Wohlstand oder Armut sein. So haben wir in ärmeren Gegenden viele Hochhäuser mit kleinen Wohnungen. Wichtig hierbei ist zu beachten, dass Personen, welche Sozialhilfe erhalten wenig finanzielle Mittel zur Verfügung haben und somit automatisch vermehrt in günstigen Gegenden mit wenig Wohnfläche pro Person vorzufinden sind. Dennoch untersuchen wir den Zusammenhang mittels 'geomap' Plot. Dieser Plot ist geeignet für Personen, welche sich geografisch im visualisierten Gebiet auch auskennen. Somit macht es keinen Sinn

diese Visualisierung

Personen zu zeigen,

welche nicht wissen,

welches Quartier wo in

Basel vorzufinden ist.

Dennoch sehen wir eine

Korrelation zwischen

Quartieren mit wenig

Wohnfläche und einem

hohen

Sozialhilfeempfängeranteil

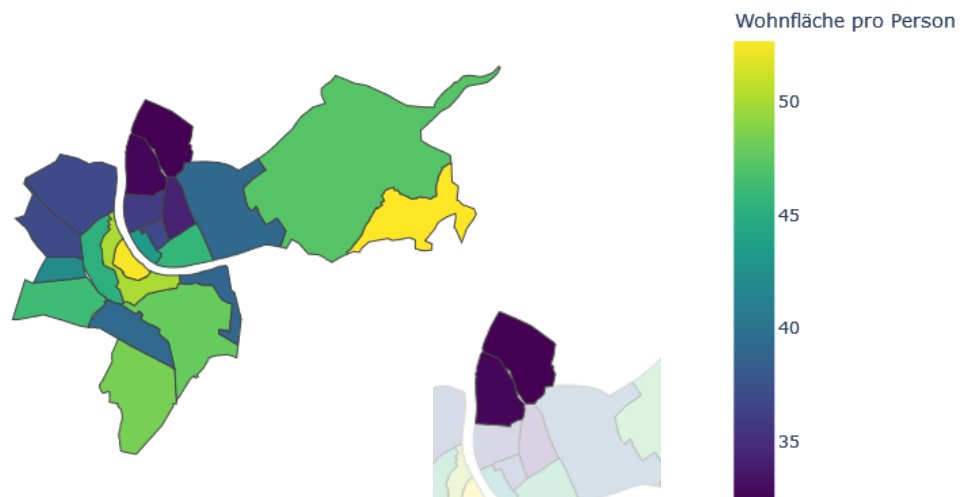
an den Beispielen

'Klybeck' und

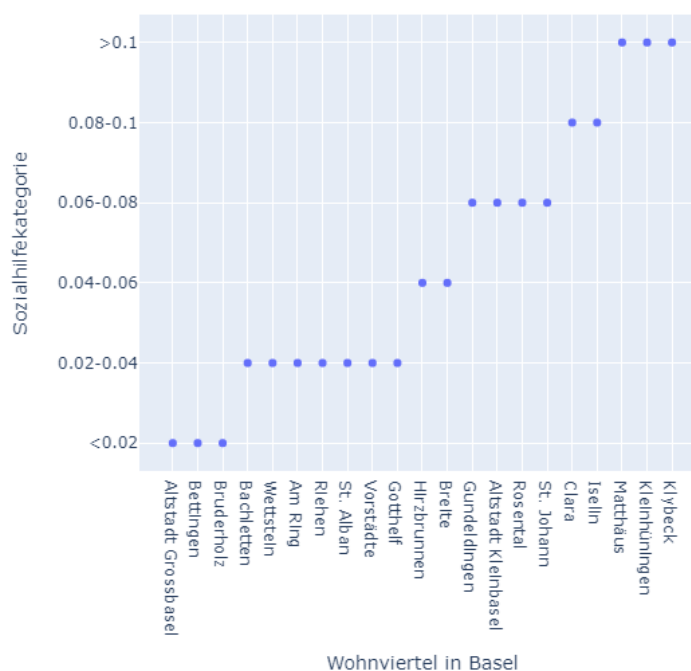
'Kleinhüningen' (Mittig im

Norden). Um dem Publikum diese Korrelationen gezielt aufzuzeigen, können die gewünschten

Quartiere markiert und hervorgehoben werden.



Verteilung Sozialhilfeempfänger Kategoriell 2021

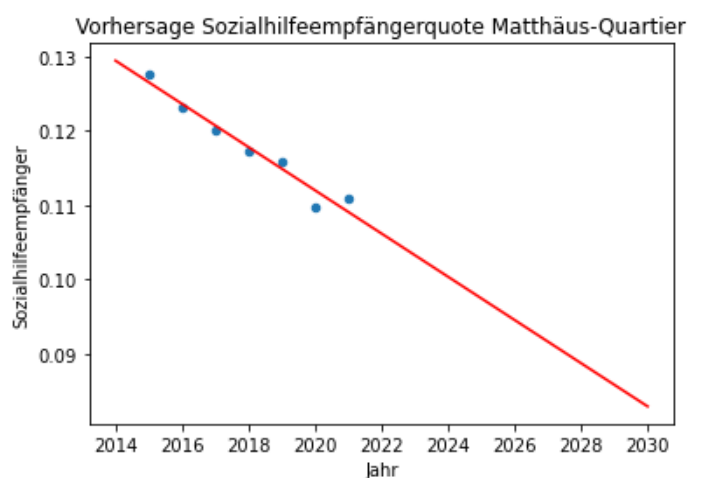
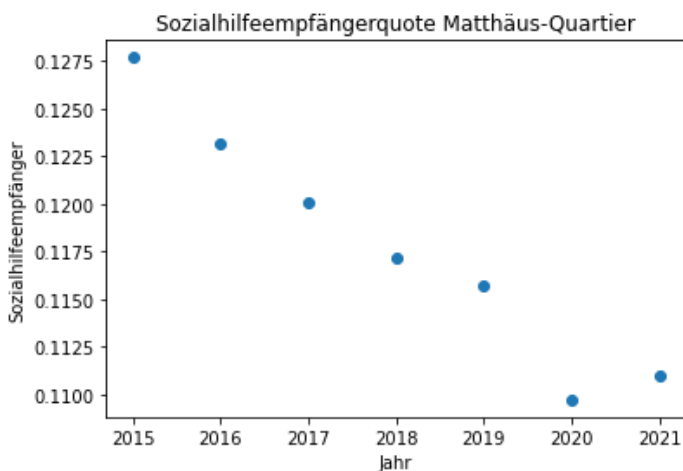
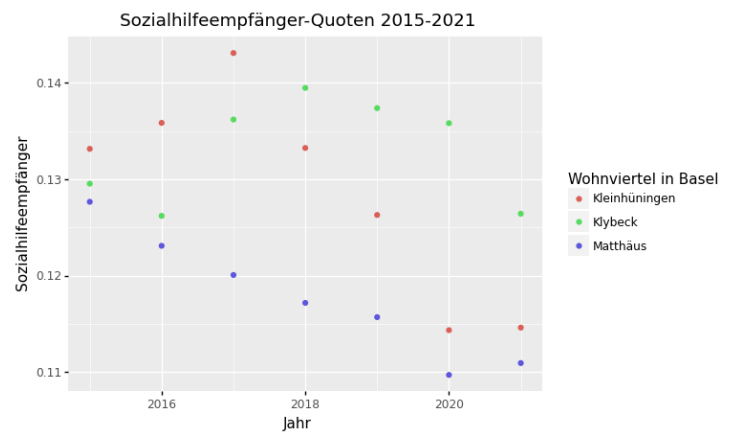


Um einen weiteren Überblick über die Anzahl Quartiere je Sozialhilfeempfängerquote-Kategorie zu erhalten, wurden die Quartiere in unterschiedliche Klassen eingeteilt. Die Abtrennungen wurden jeweils in 2% Stufen unterteilt. Der Leser sieht eine klare Abtrennung, wie viele Quartiere sich in welcher Kategorie befinden. Dies wäre vorteilhaft, wenn beispielsweise finanzielle Mittel in Form von Quartiers-Subventionen, je Kategorie ausgeschüttet werden.

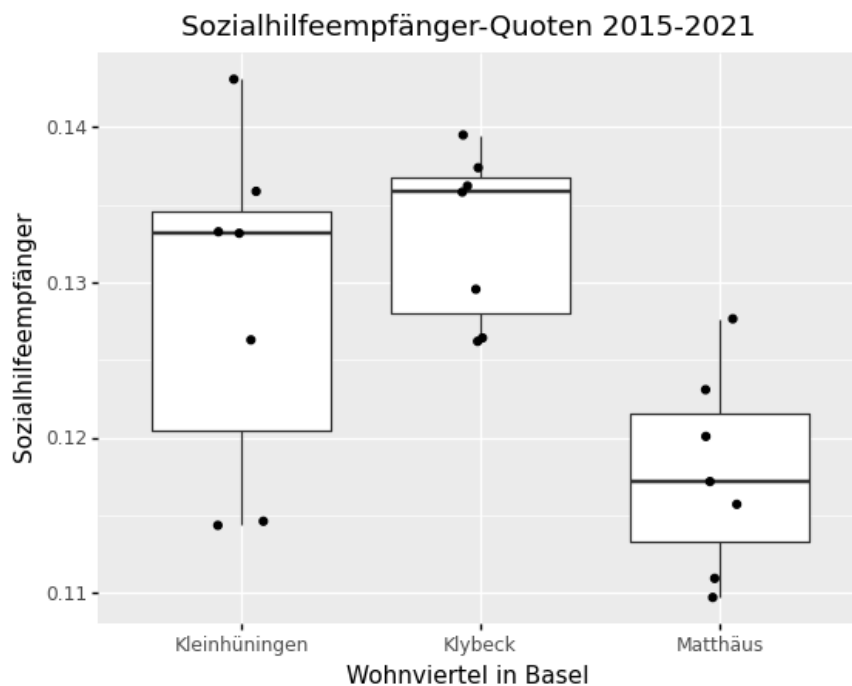
LE 4: Grammar of Graphics Tools

Die 'Grammar of Graphics Tools' ist etwas grundlegendes für gute Visualisierungen. Ähnlich wie bei der LE 2 'Visual perception' achten wir hierbei besonders auf das Publikum und wie wir die Daten darstellen. Die Komponenten, welche wir hierbei besonders beachten sind die Daten, die Ästhetik, die Skala, unser Visualisierungstyp, statistische aussagen, Fassetten sowie das Koordinatensystem [16,17]. In diesem Abschnitt werden die drei Wohnquartiere, welche die meisten Sozialhilfeempfänger im Jahre 2021 hatten, genauer untersucht. Die Visualisierungen wurden mit der Library 'Plotnine' erstellt.

Ein Vergleich der 3 Quartiere mit den grössten Sozialhilfeempfängern, zeigt, dass die Ästhetik einen grossen Einfluss auf das Visuelle Empfinden der Betrachter hat. Dieser Plot wurde mittels eines einfachen Befehles rasch geplottet. Wir sehen drei Variablen, welche mit einer guten Skalierung dargestellt werden. Ausserdem stimmt die Ästhetik. Es wurden drei sehr Verschiedene Farben gewählt, die Achsentitel werden gross und Fett beschrieben und wir erhalten alle nötigen Informationen. Jedoch würde ein anderer Diagramm-typ wie ein Linien Diagramm mehr Klarheit schaffen



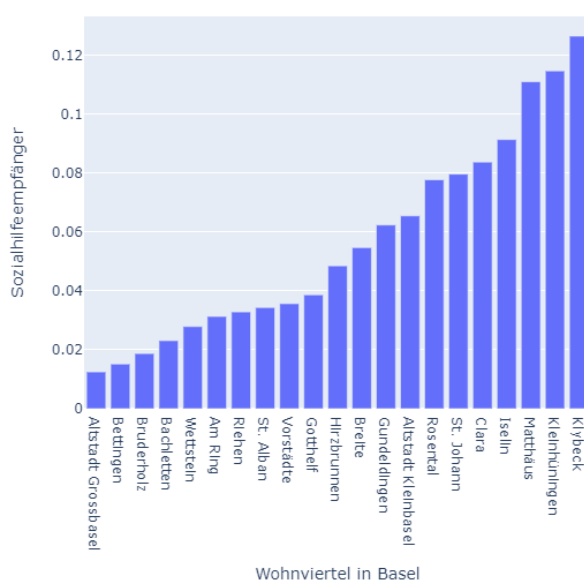
In den obenstehenden Grafiken sehen wir links, die Verteilung der Sozialhilfeempfänger der letzten sieben Jahre im Quartier Matthäus. Rechts sehen wir das Modell für die Vorhersage der Sozialhilfeempfänger-Quote vom Jahr 2014 bis zum Jahre 2030. Mithilfe einer roten Linie ist einfach verständlich, wo sich die Sozialhilfe-Quoten in den nächsten Jahren befinden werden. Hierbei ist wichtig, dass keine genaue Aussage wegen zu wenigen Datenpunkten gemacht werden kann. Das Modell weist zwar einen Score von über 95% (sehr Gut) auf, damit sehen wir jedoch nur, dass die Daten schön linear verteilt sind. Zudem ist auch klar, dass die Sozialhilfeempfänger-Quote nicht ewig linear sinkt, denn so hätten wir in einigen Jahren eine negative Sozialhilfeempfänger-Quote.



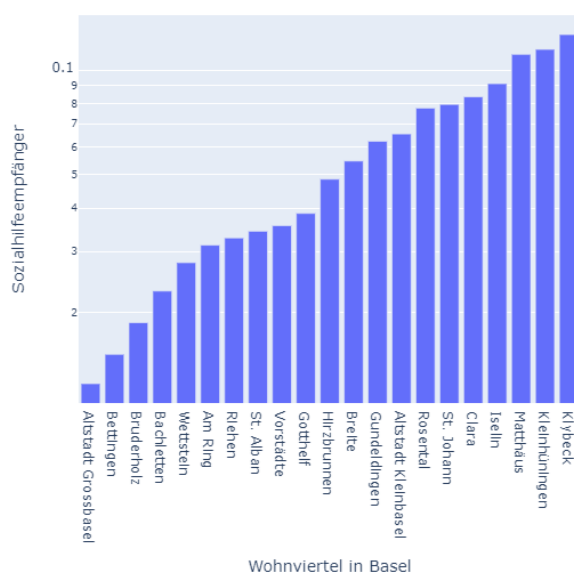
Ein weiteres wichtiges Prinzip bei Visualisierungen sind Statistische Merkmale, wie der Durchschnitt, Quantile oder auch der Median. Viele dieser Statistischen Werte können mittels Boxplot dargestellt werden. Beispielsweise sehen wir in der Box die Mittleren 50% der dargestellten Daten. Der Strich zeigt unseren Median. Die Whisker zeigen uns die Spannweite. Diese sind auf maximal 1.5 der Länge der Box beschränkt. Befänden sich Werte ausserhalb der Whisker wären diese als Ausreisser zu betrachten [18,19].

Ein weiterer wichtiger Punkt, ist die Wahl des Koordinatensystems. Dabei stehen uns zwei Varianten zu Verfügung. Das Polare Koordinatensystem, welches wir bei Diagrammen wie dem Pie- oder Donutchart antreffen. Der Vorteil dieser Diagramme wie das einfache Lesen, ohne viel beachten zu müssen bringt leider auch seine Nachteile mit sich. So ist es schwieriger für das menschliche Auge Differenzen anhand von Winkelveränderungen zu sehen (LE1 Piechart). Hingegen beim Kartesischen Koordinatensystem können wir kleine Veränderungen leicht erkennen. Dabei ist es wichtig die Achsenbeschriftung genau zu beachten. So können Veränderungen leicht durch Logarithmierte oder Abgeschnittene Achsen zum jeweiligen Zweck beeinflusst werden.

Verteilung Sozialhilfeempfänger 2021



Verteilung Sozialhilfeempfänger 2021 Logarithmische y-Achse

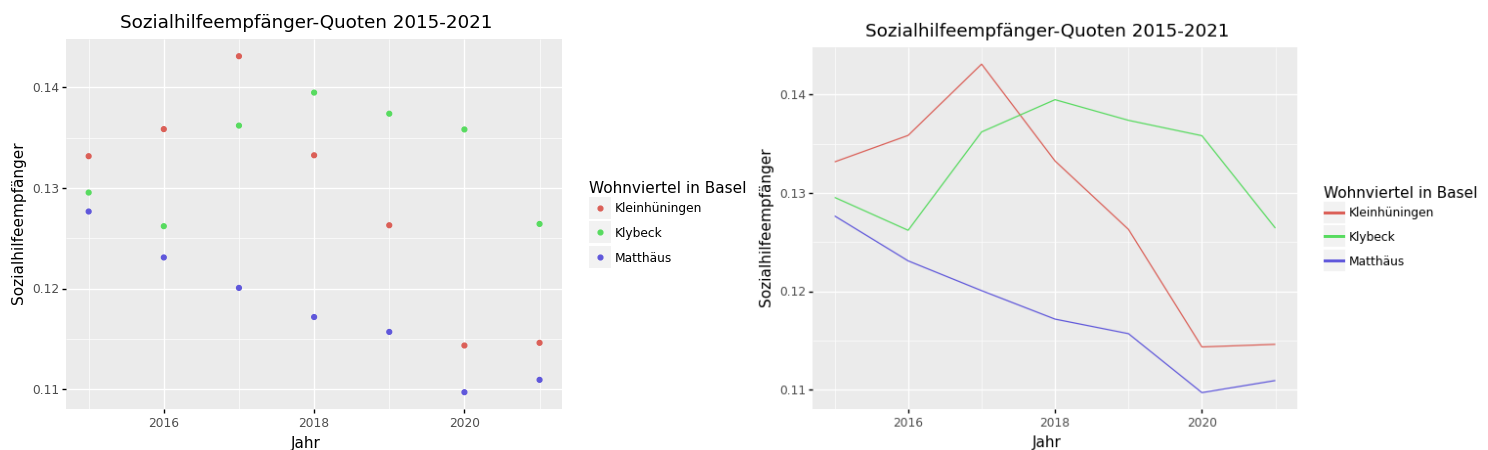


LE 5 Evaluation:

In den letzten vier Kapiteln wurden verschiedene Untersuchungen, rund um den Zusammenhang zwischen einzelnen Wohnquartieren und deren Sozialhilfeempfänger-Quote, gemacht. Dabei wurden verschiedene Daten visualisiert. Nun geht es im letzten Kapitel darum, diese Visualisierungen mittels einer kurzen Studie, auf deren Aussagekraft zu überprüfen. Es gibt verschiedene Arten von Evaluationsmethoden wie Usability-Tests, Feldstudien, Prototyping etc. dabei unterscheiden wir zwischen einem technik- oder benutzerorientierten Projekt. Ich untersuche die Wirksamkeit bestimmter Visualisierung im direkten Vergleich (technikorientiert) [22]. Das technikorientierte Experiment besteht aus den fünf folgenden Schritten: Hypothese, unabhängige Variablen, Kontrollumgebung, abhängige Variablen und Ergebnisse. Dabei ist wichtig zu beachten, dass aufgrund der geringen Teilnehmer Anzahl keine Signifikante Aussage getroffen werden kann [23].

Die Hypothese lautet: 'Die Personen können mithilfe der Visualisierung 2 (rechtsunten) die Fragestellung über das Verhalten der Sozialhilfeempfänger-Quote, schneller beantworten als mit der Visualisierung 1 (links unten). Dies Aufgrund der zusammengefassten Punkte.

H0: Es besteht kein Zeitunterschied.



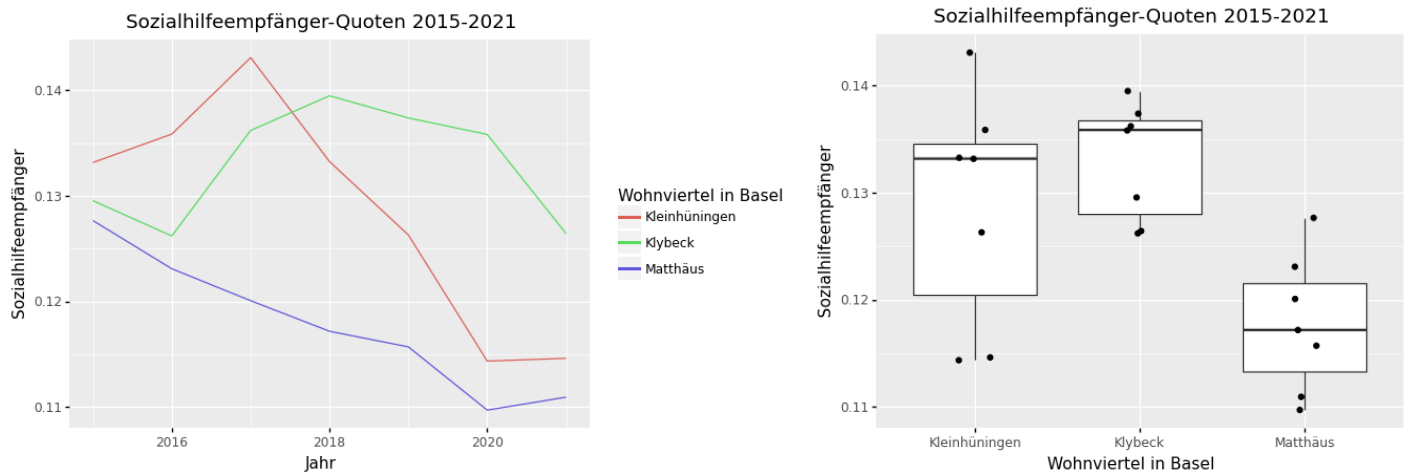
Bei dem Experiment wurden vier verschiedene Personen in jeweils zwei Gruppen unterteilt (between subject experiment). In den Gruppen waren jeweils eine männliche und eine weibliche Person. Die Personen befinden sich allesamt in unterschiedlichen Bachelor-Studiengängen (Mathematik, Geschichte, Pädagogik und Maschinenbau) und sind zwischen 21 und 25 Jahren alt. Die Visualisierungen, welche dieselben Daten abbildet, wurden je einer Versuchsgruppe, mit der folgenden Fragestellung gezeigt: 'Welche Sozialhilfeempfänger-Quote veränderte sich in den letzten Jahren am meisten?'. Nebenbei wurde die Zeit bis zu einer korrekten Antwort gestoppt.

Das Experiment wurde bei mir zuhause in meinem Büro unter Ausschluss des anderen Teams durchgeführt.

Resultate:

Der Zeitunterschied war mit der mehr als doppelten Zeit sehr gross (12,44/5,38; 10,81/5,07:). Ein Line Plot ist somit besser geeignet, um Daten über einen Zeitraum abzubilden und den Verlauf abzulesen. Zudem kam bei der Versuchsgruppe 1 (Scatter Plot) der Wunsch auf die Punkte zu verbinden.

Nach einer Pause wurde ein zweiter Versuch mit denselben Versuchs-Paaren durchgeführt. Nun erhielt Gruppe Scatterplot den Lineplot und Gruppe Lineplot den Boxplot. Die Versuchsumgebung ist wieder in meinem Büro, die Fragestellung jedoch leicht verändert. Zudem sind die X-Achsen nicht mehr genau dieselben, was somit bedeutet, dass nicht mehr in beiden Diagrammen dieselben Daten dargestellt werden. Die neue Fragestellung lautet: 'Welches Quartier hat die kleinste Spannweite über die Jahre der Sozialhilfeempfänger-Quote.' Meine neue Hypothese lautet: 'Mithilfe des Boxplots Diagramm ist die Spannweite einfacher abzulesen als jene mit einem Line Plot.'

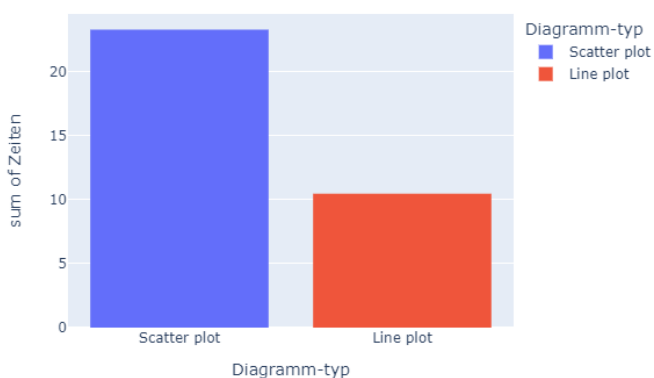


Resultat:

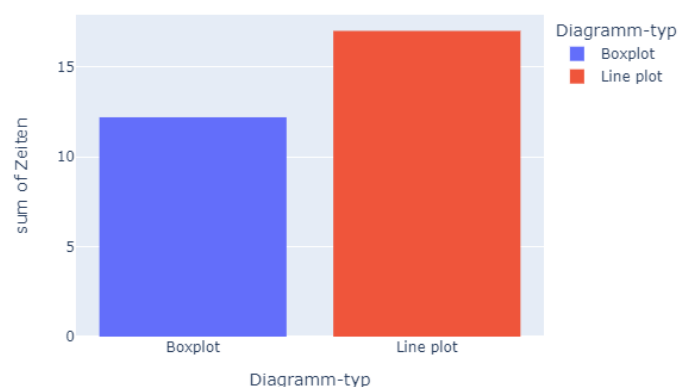
Mit einem durchschnittlichen Zeitunterschied von rund zwei Sekunden (7,65/5,28;9,34/6,91) kann lediglich ein kleiner Unterschied festgestellt werden, welcher man nun mit einer grösseren Versuchsgruppe genauer untersuchen müsste. jedoch können wir vermuten, dass Spannweiten wie angenommen mit einem Boxplot besser abzulesen sind. Zudem muss auch beachtet werden, dass die Versuchsteilnehmer bereits mit den Daten und einem, auch wenn nicht gleichen Diagramm bekannt waren. Der Zeitunterschied wäre bei diesem fehlenden Wissen eventuell grösser ausgefallen. Eine Erkenntnis aus den Untersuchungen ist, dass jedes Diagramm auf die Aufgabe zugeschnitten sein sollte, so war ersichtlich, dass das Linien-Diagramm für eine Aufgabe besser und für eine andere Aufgabe weniger gut geeignet ist.

Visualisierte Auswertungen:

Vergleich 1 Versuchsgruppen Zeiten



Vergleich 2 Versuchsgruppen Zeiten



Anhang

Datensatz

Die Diagramme wurden alles mittels einem Datensatz, zur Verfügung gestellt vom Kanton Baselstadt, erstellt.

<https://opendata.swiss/de/dataset/kennzahlen-zu-den-basler-wohnvierteln-und-landgemeinden>

Quellen

Allgemein

- [1] <https://www.bfs.admin.ch/bfs/de/home/statistiken/soziale-sicherheit/sozialhilfe/sozialhilfebeziehende.html>
- [2] <https://www.bfs.admin.ch/bfs/de/home/statistiken/soziale-sicherheit/sozialhilfe/sozialhilfebeziehende/wirtschaftliche-sozialhilfe.html>
- [3] <https://plotly.com/python/>
- [4] <https://plotnine.readthedocs.io/en/stable/index.html>

LE1

- [5] <https://www.storytellingwithdata.com/blog/2013/04/chart-chooser>
- [6] <https://visme.co/blog/types-of-graphs/>
- [7] <https://www.youtube.com/watch?v=C07k0euBpr8>
- [8] <https://www.youtube.com/watch?v=qGalB-bRn-A>

LE2

- [9] <https://spaces.technik.fhnw.ch/multimediathek/video/the-beauty-of-data-visualization-david-mccandless>
- [10] <https://www.youtube.com/watch?v=0Smgm2UTUSo>
- [11] <https://www.youtube.com/watch?v=4pymfPHQ6SA>

LE3

- [12] <https://www.eea.europa.eu/data-and-maps/daviz/learn-more/prepare-data>
- [13] <https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4>
- [14] <https://developer.ibm.com/tutorials/ba-cleanse-process-visualize-data-set-1/>
- [15] <https://realpython.com/python-data-cleaning-numpy-pandas/>

LE4

- [16] <https://towardsdatascience.com/a-comprehensive-guide-to-the-grammar-of-graphics-for-effective-visualization-of-multi-dimensional-1f92b4ed4149>
- [17] <https://www.youtube.com/watch?v=kepKM7Z2O54>
- [18] <https://chartio.com/learn/charts/violin-plot-complete-guide/>
- [19] <https://datacarpentry.org/python-ecology-lesson/07-visualization-ggplot-python/index.html>
- [20] <https://datacarpentry.org/python-ecology-lesson/07-visualization-ggplot-python/index.html>

LE5

- [21] <https://spaces.technik.fhnw.ch/multimediathek/video/usability-testing-w-5-users-design-process-video-1-of-3>
- [22] <https://www.nngroup.com/articles/usability-testing-101/>
- [23] <https://novustat.com/statistik-glossar/hypothesentest.html>