1. JUNI 2020

Datensatz-orientierte, automatische Auswahl raum-zeitlicher Visualisierungen

Eine nutzerfreundliche und effiziente Visualisierung offener raumzeitlicher Daten der mCLOUD

Zepner Laura[a], Jirka Simon[b], Schulte Jan[b], Sauer Petra[c], Darabisahneh Bayan[c], Possienka Rico[c], Mäs Stephan[a]

[a] Professur für Geoinformatik, Technische Universität Dresden
[b] 52° North Initiative for Geospatial Open Source Software GmbH
[c] Professur Informatik, Beuth Hochschule für Technik Berlin

Inhaltsverzeichnis

1 Einleitung

Das Angebot frei zugänglicher Geodaten der Verwaltungen wächst rasant. Wie kann ein Nutzer möglichst schnell erkennen, ob ein oder mehrere Datensätze für seinen Anwendungsfall relevant sind und geeignete (Geo-)Informationen liefern können? Die Bereitstellung möglichst einfacher, intuitiver Visualisierung der Daten zur effizienten und nutzerfreundlichen Exploration ist hier essentiell. Allerdings ist eine schnelle und informative Visualisierung der offenen Daten aus der mCLOUD derzeit noch schwierig, insbesondere für raumzeitliche Daten, die i. d. R. zunächst heruntergeladen und in gängige Datenformate konvertiert werden müssen, bevor eine visuelle Exploration der Daten und Analyse des Informationsgehalts möglich ist. Dadurch sind oft nur GI-Experten qualifiziert genug die Exploration raumzeitlicher, offener Daten durchzuführen.

Aktuell existieren vereinzelt spezielle Anwendungen (wie zum Beispiel das Open Sensor Web[1] oder die Visualisation Sandbox[2] des EU Open Data Portal) mit denen eine Exploration für Laien ermöglicht wird. Die Auswahl ist jedoch gering und auch hier ist oft fundiertes Fachwissen notwendig um geeignete Visualisierungen für einen spezifischen Datensatz auszuwählen.

1.1. Über das Projekt

mVIZ zielt auf die nutzerfreundliche Visualisierung offener raumzeitlicher Daten. Dazu wurde in einer Vorstudie eine Methodik entwickelt, die die Auswahl und Erstellung nutzerfreundlicher Visualisierungen für offene raumzeitliche Daten der mCLOUD mit Fokus auf Usability und Nachnutzung unterstützt.

Der vorliegende Leitfaden beschreibt die Methodik und dient als Grundlage zur Konzeption, Erweiterung oder Verbesserung von Visualisierungswerkzeugen bzw. zu deren Weiterentwicklung und Integration in offene Datenportale.

Abbildung 1 beschreibt die Teilziele der Vorstudie. Davon werden folgende Aspekte in diesem Leitfaden ausführlicher besprochen:

- Eine Bestandsaufnahme offener raumzeitlicher Daten in der mCLOUD sowie eine Übersicht über verfügbare raumzeitliche Visualisierungen und Analysewerkzeuge (siehe 2.2 Analyse der Datensätze)
- Die Methodik zur Auswahl geeigneter Visualisierungen für raumzeitliche Daten
- Der Demonstrator für ausgewählte Daten der mCLOUD und Visualisierungen sowie dessen Evaluierung

[1] https://www.opensensorweb.de

[2] https://data.europa.eu/euodp/en/node/6072

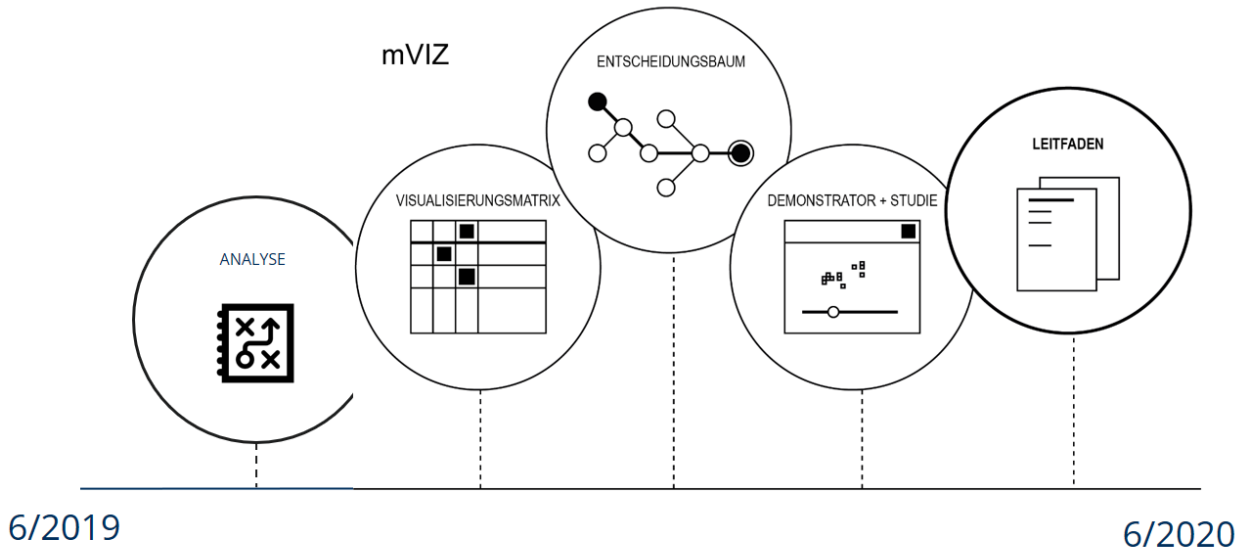


Abb. 1: Teilaspekte des mVIZ Projektes

Damit sollen Entscheidungsunterstützungen bei der Aufbereitung der Daten und Metadaten, Datenbereitstellung bis hin zur Erstellung der Visualisierung geliefert werden. Hinweise zu Anforderungen an die Metadaten und Datenpublikationen, sowie eine Auswahl passender Visualisierungsmethoden abhängig von zu explorierenden Daten sollen es ermöglichen im späteren Verlauf eine automatische Auswahl nutzerfreundlicher Visualisierungen für die interaktive Exploration offener raumzeitlicher Daten zu treffen.

1.2. Die Zielgruppe

Das übergeordnete Ziel von mVIZ ist es, Nutzer von Open-Data-Portalen bei der Auswahl von Datensätzen zu unterstützen. Somit werden diese als übergeordnete Zielgruppe für das Projekt betrachtet. Ihre Bedürfnisse wurden in mehreren Umfragen analysiert, welche darauf abzielten, welche Visualisierungen aktuell genutzt werden und welche Rolle Open-Data-Portale und deren Funktionen spielen.

Dieser Leitfaden richtet sich jedoch an Datenproduzenten und Anwendungsentwickler. Er bietet eine strukturierte Unterstützung bei der Aufbereitung der Daten und Metadaten bis hin zur Erstellung der Visualisierung. Dazu soll der Leitfaden Entscheidungspunkte liefern, die es ermöglichen eine automatische Auswahl von Visualisierungen für einen Datensatz zu treffen.

Anwendungsentwicklern soll ein fundiertes Verständnis davon vermittelt werden, welche Kriterien für die automatische Auswahl einer Visualisierung notwendig sind und wie diese in Systeme integriert werden können. Des Weiteren sollen Ideen zur Softwareentwicklung, z.B. hinsichtlich Architektur und verfügbarer Bibliotheken vorgestellt werden.

Für Datenproduzenten soll der Leitfaden Aufschluss über notwendige Aspekte zur Datenaufbereitung geben, wie z.B. über mögliche Datenstrukturen, -formate und

Minimalanforderungen an Datenbeschreibungen. Der Leitfaden gibt einen Überblick über die aktuelle Verwendung von Metadaten und analysiert, welche Angaben fehlen zu einer erfolgreichen automatischen Auswahl von Visualisierungen. Datenproduzenten sollen am Ende in der Lage sein zu entscheiden, wie Daten beschrieben werden können und welche Datenformate sich besonders gut eignen.

2 Stand der Technik

2.1. Konzept

Um ein sinnvolles Konzept zu erstellen wurde zunächst eine Bestandsanalyse der Teilaspekte durchgeführt. Diese gliedert sich in die folgenden drei Bereiche:

- Analyse aktueller raumzeitlicher Datensätze in der mCLOUD
- Analyse aktueller raumzeitlicher Visualisierungen
- Analyse der Nutzungsmerkmale (Kontext, Nutzer, Anforderungen)

Aus den Ergebnissen dieser Analysen entsteht die Visualisierungsmatrix. Wie in Abb. 2 zu sehen ist, wird hierfür eine Zuordnung von Visualisierungen und Nutzungsmerkmalen zu den Merkmalen der Daten durchgeführt.

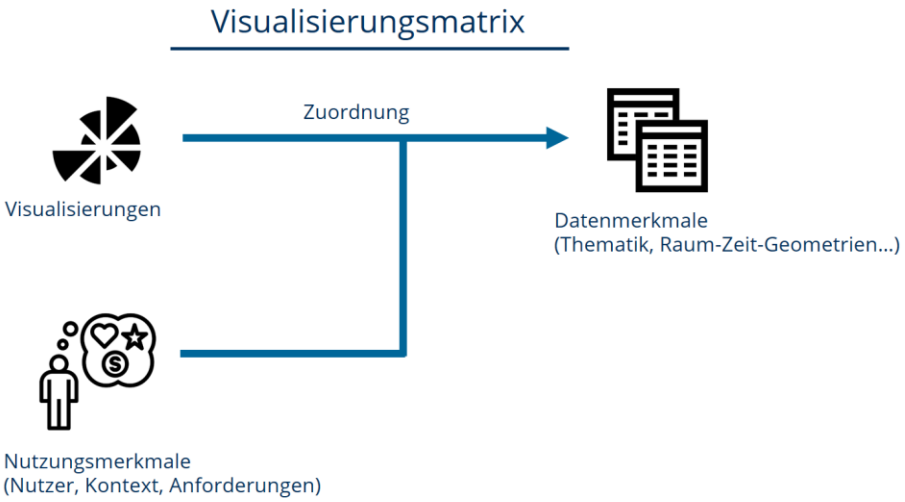


Abb. 2: Konzept für die Visualisierungsmatrix (Quellen siehe: Icon Referenzen)

Durch die Klassifizierung sämtlicher Teilaspekte ist es möglich eine gemeinsame Sprache zwischen Datenmerkmalen, Visualisierung und Nutzungsmerkmalen zu erschaffen und diese als Grundlage für ein kohärentes Gesamtkonzept zu nutzen.

Aus der dadurch entstanden Visualisierungsmatrix können anschließend, in Kombination mit diversen Nutzungsmerkmalen (Kontextabhängige Anforderungen), Entscheidungspunkte extrahiert werden (siehe Abb. 3). Basierend auf der Klassifikation der Visualisierungen und dem

dadurch zugrundeliegenden Empfehlungskatalog kann eine automatische Auswahl der geeigneten Visualisierung getroffen werden.

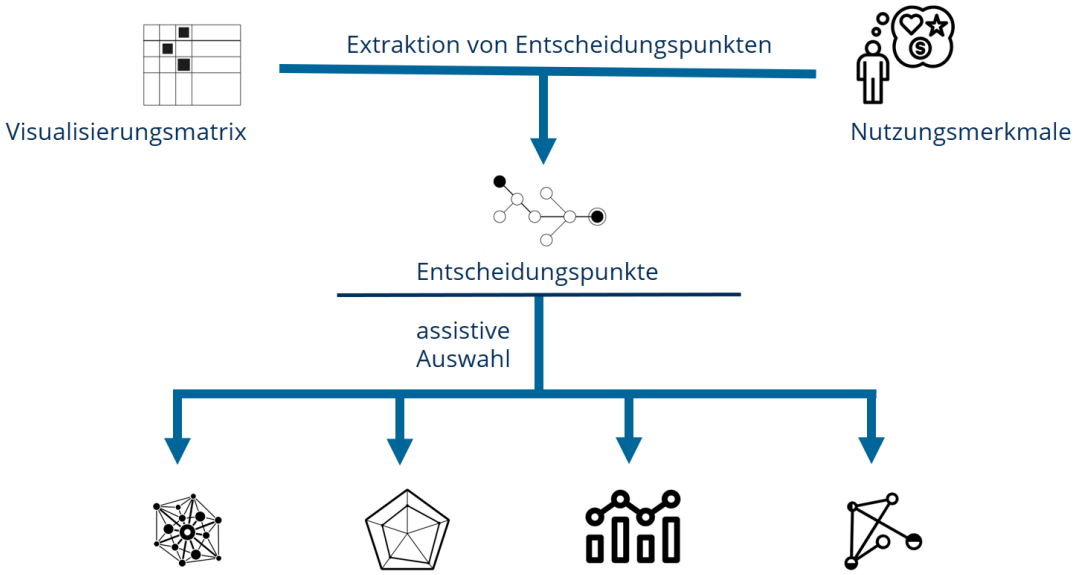


Abb. 3: Konzept zur automatischen Auswahl geeigneter Visualisierungen (Quellen siehe: Icon Referenzen)

Das Ergebnis ist eine kleine Liste verschiedener Visualisierungen, welche zum ausgewählten Datensatz und den Nutzungsmerkmalen passen.

2.2. Analyse der Datensätze

Die Analyse der Datensätze in Form der zum Analysezeitpunkt verfügbaren Datenpublikationen der mCLOUD erfolgte einerseits hinsichtlich allgemeiner Kriterien wie z.B. Kategorien, Publikationszeitpunkt, Anbieter oder Datenformate bzw. Zugriffsart. Andererseits erfolgte eine Analyse ausgewählter Datenpublikationen hinsichtlich raumzeitlicher Merkmale, die sowohl aus den Daten selbst als auch aus den Metadaten gewonnen wurden. Nachfolgend werden beide Analysen nacheinander vorgestellt.

Analyse nach allgemeinen Kriterien

Zum Zeitpunkt der Analyse im November 2019 befanden sich insgesamt 1.526 Datenpublikationen in der mCLOUD. Im Vergleich zur ersten punktuellen Analyse zum Zeitpunkt des Projektstarts im Juni 2019 mit 1.141 Datenpublikationen ist das eine Zunahme von 385 Datenpublikationen bzw. ca. 33% innerhalb von vier Monaten, was auf einen regen Gebrauch zumindest des Angebots der Bereitstellung von Daten schließen lässt. Zum Zeitpunkt des Projektendes im Mai 2020 sind 1.910 Datenpublikationen in der mCLOUD verfügbar, die eine Gesamtanzahl von 6.322 Dateien umfassen. Innerhalb von 11 Monaten sind somit knapp 800 neue Datenpublikationen auf die mCLOUD gestellt worden, was gut aufzeigt, wie positiv sich der

bereits bis 2019 zu konstatierende Trend der Bereitschaft zur Datenbereitstellung weiterentwickelt.

Allgemeine Kriterien: Kategorien

Die Datenpublikationen der mCLOUD sind in sechs Kategorien unterteilt, jede Datenpublikation kann jedoch mehreren Kategorien zugeordnet sein. Die Kategorien sind: „Wasserstraßen und Gewässer“, „Straßen“, „Klima und Wetter“, „Bahn“, „Infrastruktur“ sowie „Luft- und Raumfahrt“ (Abb. 4). Mit 631 Datenpublikationen ist „Wasserstraßen und Gewässer“ die umfangreichste Kategorie, gefolgt von „Straßen“ mit 485 Datenpublikationen und „Klima und Wetter“ mit 291. Die geringste Anzahl an Datenpublikationen ist der Kategorie „Infrastruktur“ mit 43 Datenpublikationen zugeordnet.

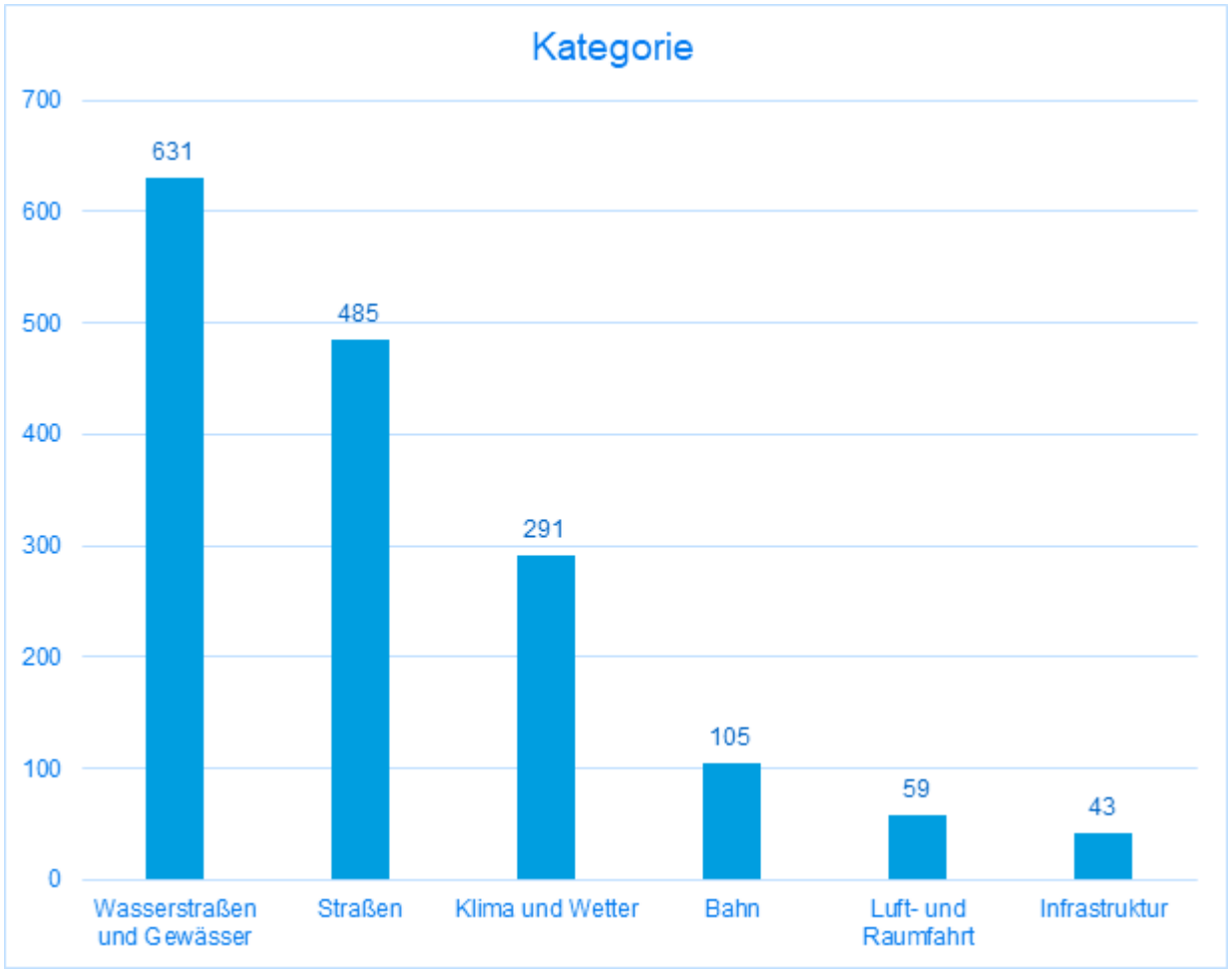


Abb. 4: mCLOUD-Datensätze nach Kategorien (Nov. 2019)

Als interessant erweist sich der Vergleich mit den Zahlen von Juni 2019. Hier war die Kategorie „Straßen“ mit 439 Datenpublikationen die stärkste Kategorie vor „Wasserstraßen und Gewässer“ mit 305 Datenpublikationen. Dort waren jeweils nur ca. die Hälfte der Datenpublikationen verfügbar, die im November 2019 bereitstehen.

Allgemeine Kriterien: Zeitpunkt der Veröffentlichung in mCLOUD

Nach dem Zeitpunkt der Veröffentlichung der Datenpublikationen in der mCLOUD ergibt sich folgendes Bild (Abb. 5):

- Die frühesten Datenbestände stammen aus dem Jahr 1994 mit drei Datenpublikationen zur Luftqualität in Berlin. Diese wurden aber erst im Jahr 2018 publiziert.
- Die früheste Datenpublikation stammt aus dem Jahre 2007. Hier wurde eine Datenpublikation aus dem Bereich „Wasserstraßen und Gewässer“ zur Verfügung gestellt.
- Nachdem über fast 10 Jahre keine signifikante Entwicklung zu verzeichnen war, erfolgte der stärkste Anstieg der Datenpublikationen im Jahr 2019 mit 974 neuen Datenpublikationen. (Dieser Trend setzt sich auch im Jahr 2020 fort. Bis Anfang Mai 2020 wurden bereits 549 neue Datenpublikationen zur Verfügung gestellt.)

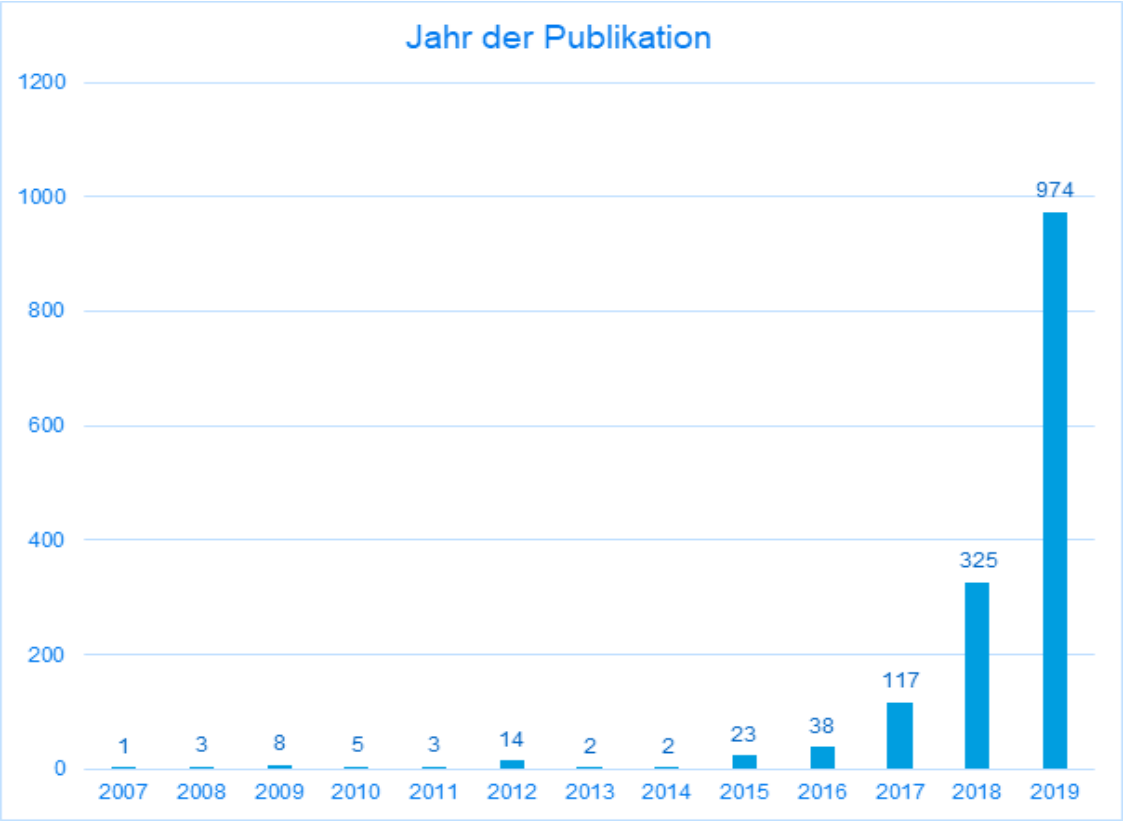


Abb. 5: mCLOUD-Datenbestände nach dem Jahr der Publikation

Allgemeine Kriterien: Anbieter

Insgesamt 121 Anbieter haben (Stand November 2019) Daten in der mCLOUD zur Verfügung gestellt (Abb. 6). Mit Abstand die meisten Datensätze (424) werden hierbei vom Bundesamt für Seeschifffahrt und Hydrographie (BSH) gestellt, was den hohen Anteil an Datenpublikationen in der Kategorie „Wasserstraßen und Gewässer" erklärt. Mit einigem Abstand auf dem zweiten Platz ist das BMVI selbst mit 283 Datenpublikationen. Insgesamt 26 Anbieter stellten zwischen zehn und 51 Datenpublikationen bereit. Die 93 verbleibenden Anbieter stellen weniger als zehn Datenpublikationen, 48 davon lediglich eine Datenpublikation zur Verfügung.

Die Liste der Anbieter setzt sich zusammen aus unterschiedlichen Institutionen und Unternehmen. Einige sind Einrichtungen oder Behörden des Bundes, des Landes (wie das Landesamt für Natur, Umwelt und Verbraucherschutz NRW) oder der Städte, wobei bei letzteren die größten Anbieter die Städte Köln, Moers und Bonn sind. Aber auch Städte wie Berlin, Rostock, Münster ziehen nach und bieten zunehmend Datenpublikationen an. Die European Space Agency (ESA) ist der einzige Anbieter auf europäischer Ebene, der Tagesspiegel das einzige Medium. Darüber hinaus haben Forschungseinrichtungen (z.B. das Leibnitz-Institut für ökologische Raumentwicklung IÖR), privatwirtschaftliche Unternehmen (hier insbesondere die smile Consult GmbH) sowie die Deutsche Bahn und deren Sparten (DB Netz AG, DB Fernverkehr AG) Daten für die mCLOUD zur Verfügung gestellt.

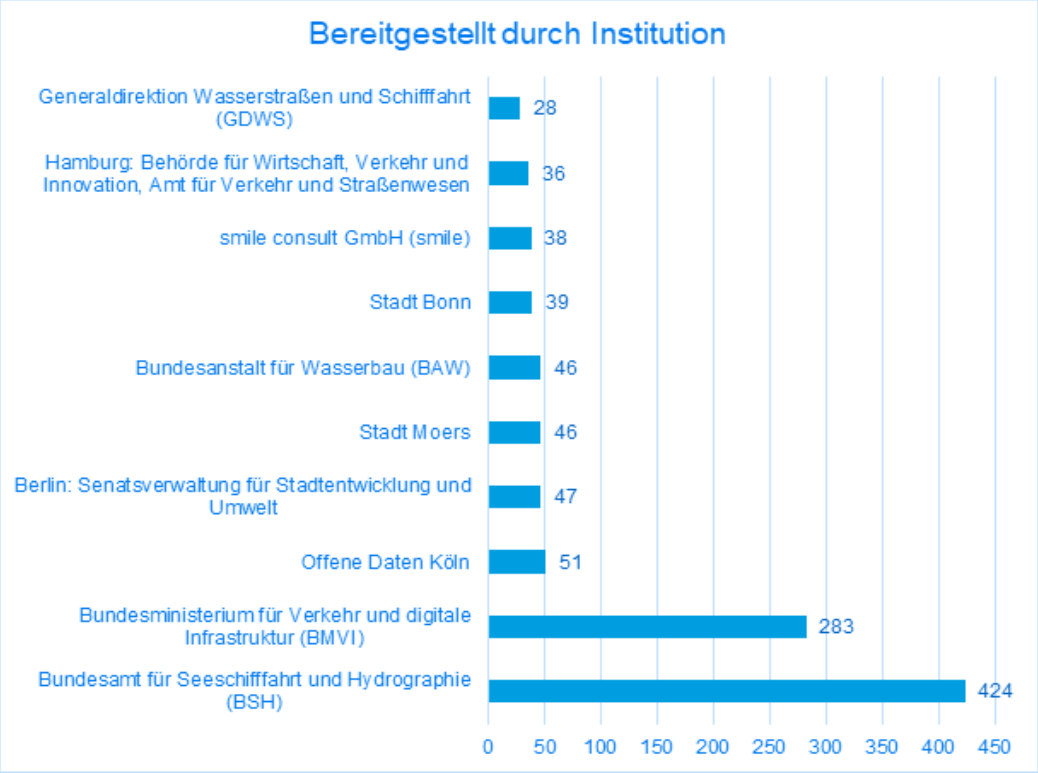


Abb. 6: mCLOUD Datenbestände nach Anbietern

Betrachtet man die Anzahl der Anbieter nach der Art sowie den von ihnen zur Verfügung gestellten Datensätzen, ergibt sich ein uneinheitliches Bild. Die meisten Anbieter sind Einrichtungen von Städten (42), mit deutlichem Abstand gefolgt von denen des Bundes (24) und privatwirtschaftlichen Unternehmen (20).

Nach Anzahl der Datenpublikationen haben Bundeseinrichtungen mit 917 Datenpublikationen die meisten Datenpublikationen zur Verfügung gestellt, gefolgt von den Städten mit 390 und den privatwirtschaftlichen Unternehmen mit 70. Die Länder nehmen mit elf Anbietern und 56 Datenpublikationen eine mittlere Position ein, die Deutsche Bahn hat mit 14 Anbietern 46 Datenpublikationen verfügbar gemacht. Forschungseinrichtungen sind mit elf Anbietern und 25 Datenpublikationen schwächer vertreten. Die ESA als einziger europäischer Anbieter hat 22 Datenpublikationen zur Verfügung gestellt, während vom Tagesspiegel nur eine Datenpublikation stammt.

Allgemeine Kriterien: Zugangsarten und Datenformate

In der mCLOUD ist es möglich, Datenpublikationen nach der Art und Weise ihrer Bereitstellung zu filtern. So werden Datenpublikationen zum Beispiel über ein Portal, einen FTP-Server, als ZIP-Download, als CSV- oder Excel-Datei, als Web Feature Service (WFS) oder als GeoJSON-Datei angeboten. Hier lässt sich bereits erkennen, dass der Filter „Zugang“ nicht nach Zugangsart und Dateiformat trennt und diese Dimensionen vermischt. Eine CSV- oder Excel-Datei kann sich auf einem FTP-Server befinden, als ZIP-Datei heruntergeladen oder über ein Portal bereitgestellt werden. Datenpublikationen in der mCLOUD können somit mehrere Einträge für diesen Filter erhalten, ähnlich dem „Kategorie“-Filter. Dies ergibt Sinn, da Daten auf unterschiedliche Weise bereitgestellt werden können, was auch durchaus wünschenswert ist. So steht als Beispiel eine Datenpublikation „Verkehrszeichen“ der Stadt Rostock als XLSX, WFS, CSV, GML, GeoJSON, KML und WMS bereit. Gleichwohl könnte man über eine Aufteilung des Filters nach „Zugangsart“ und „Dateiformat“ nachdenken. Etwas erschwerend kommt hinzu, dass der Filter eine ODER-Verknüpfung umsetzt. Wählt man z.B. „Portal“ und “WFS“ aus, erscheinen nicht zwingend nur Datenpublikationen der Schnittmenge, die beide Möglichkeiten gleichermaßen anbieten.

Nach Datenformat ist der Spitzenreiter das Format „Portal“ mit 491 Datenpublikationen, gefolgt von Dateidownload mit 262 und FTP mit 257 Datenpublikationen.

Wenn man die Datenpublikationen nach Datenformaten betrachtet, ist Vorsicht geboten, da einige der erwähnten Zugangsarten die Datenformate implizieren. So liefert ein WFS in der Regel Daten in GML, einige AtomFeeds spezifizieren das Dateiformat und ein WMS ist Zugangsart und Dateiformat gleichermaßen. GML basiert wiederum auf XML. Die Liste der meistgenutzten Datenformate führt der WMS mit 253 Datenpublikationen an. Hier ist allerdings anzumerken, dass ein WMS im Regelfall nur als eine von mehreren Möglichkeiten angeboten wird. Somit kann der Interessent wählen, ob er die Rohdaten oder die Kartendarstellung mittels WMS geliefert bekommen möchte. Auf den Plätzen zwei und drei folgen die Datenformate ASCII (über

AtomFeed) mit 237 und GML (über WFS) mit 190 Einträgen. JSON- und GeoJSON-Dateien werden nur selten angeboten, mit 57 bzw. 38 Datenpublikationen.

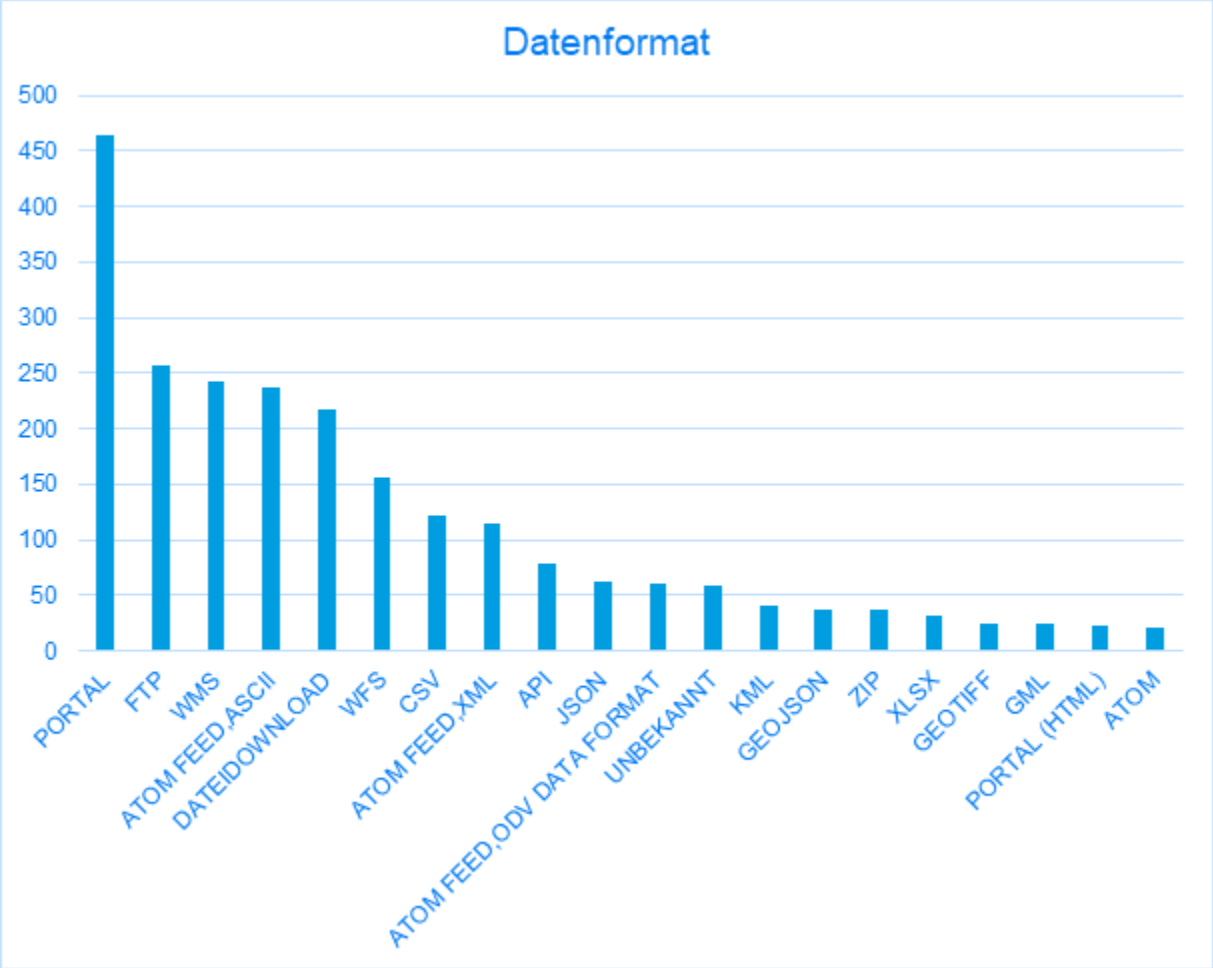


Abb. 7: mCLOUD-Datenbestände nach Datenformat

Allgemeine Kriterien: Nutzungslizenz

Die mit Abstand am häufigsten vorzufindende Lizenzart ist „Datenlizenz Deutschland Namensnennung“ (Tabelle 1). Diese Lizenz setzt keine engen Schranken und erlaubt die „kommerzielle und nicht kommerzielle Nutzung“, dabei insbesondere auch die Vervielfältigung, Veränderung, Bearbeitung und Übermittlung an Dritte, die Zusammenführung mit Daten anderer Quellen und die Einbindung in „Anwendungen in öffentlichen und nicht öffentlichen elektronischen Netzwerken“. Hierbei ist sicherzustellen, dass der Quellenvermerk die Bezeichnung des Bereitstellers, den Vermerk auf die Lizenz mit Verweis auf den Lizenztext sowie einen Verweis auf die Datenpublikation (URI) enthält. Wurden Daten verändert und bearbeitet, ist dies mit Hinweis zu versehen. Mit 363 Datenpublikationen stehen an zweiter Stelle die „Nutzungsbestimmungen für die Bereitstellung von Geodaten des Bundes“. Auf dem dritten Platz

mit 187 folgt die Creative Commons Lizenz CC-BY, die lediglich die Namensnennung vorsieht, aber sowohl kommerzielle als auch nicht-kommerzielle Nutzung, Veränderung, Bearbeitung und Weitergabe erlaubt. Nach den Datenpublikationen mit unbekannten Nutzungsbestimmungen folgen als Lizenzart noch zwei verschiedene Zero-Lizenzen, die keinerlei Bedingungen und Urheberrecht vorsehen und die Daten somit der Public Domain zuordnen lassen.

Lizenz	Anzahl
Datenlizenz Deutschland Namensnennung	630
Nutzungsbestimmungen für die Bereitstellung von Geodaten des Bundes	363
Creative Commons Namensnennung – 4.0 International (CC BY)	187
Unbekannt	138
Datenlizenz Deutschland – Zero – Version	76
Creative CC Zero License (cc-zero)	62
Nutzungsbestimmungen für die Bereitstellung von Geodaten des Landes Berlin	47

Tabelle 1: mCLOUD-Datenbestände nach Lizenzart

Analyse nach raumzeitlichen Merkmalen

Während einige der Kategorien, die in der mCLOUD ein Filtern ermöglichen, für die Analyse der raumzeitlichen Charakteristika herangezogen werden konnten, sind Kategorien wie Lizenzierung oder Anbieter eher von geringer Bedeutung und werden im Folgenden nicht mehr aufgegriffen.

Nach den Dimensionen Raum, Zeit, Thematik und Meta, wobei unter letzterer vor allem die verfügbaren Metadaten zusammengefasst werden sollten, wurden ausgewählte Datenpublikationen der mCLOUD weitergehend analysiert. Für die Analyse wurde eine Merkmalsmatrix, wie sie in Tabelle 2 als Auszug ersichtlich ist, entwickelt. Diese enthält die Merkmale der Dimensionen sowie konkrete Ausprägungen von ausgewählten Datenpublikationen der mCLOUD.

In der Dimension Raum wurde hinsichtlich der Merkmale unterschieden nach:

- der Art des Objekttyps,
- den Objekttyp-Details und
- der räumlichen Ausdehnung.

In der Dimension Zeit wurde die Analyse nach den folgenden Merkmalen und ihren Ausprägungen realisiert:

- der Erhebungszeit als Zeitangabe,
- der konkreten Zeitangabe,
- der Klassifikation nach dem Enthaltensein der Zeitangabe in Metadaten, Dateiname etc.,
- der Genauigkeit,
- der Auflösung,
- der Gültigkeit,
- der Ordnung, wobei hier nach linear, verzweigend und zyklisch unterschieden wurde.

Die thematische Dimension wurde anhand der folgenden Merkmale analysiert:

- zunächst nach der Domäne, die bereits oben als Kategorie angesprochen wurde und den Kategorien der mCLOUD folgt,
- nach dem genutzten Vokabular (Standard- vs. proprietäres Vokabular),
- nach dem Vorhandensein einer Klassifikation,
- nach der Anzahl an Themen, wobei nur zwischen ein- und mehrdimensional unterschieden wurde,
- nach dem Skalenniveau sowie
- nach der Art der Erhebung bzw. Berechnung.

Die Dimension Meta berücksichtigte einen Teil der oben bereits analysierten Kategorien wie

- Format (Daten- und Zugriffsformat),
- Quelle der Daten sowie
- Lizenz.

Ausgewählte Datenpublikationen mit raumzeitlichen Merkmalen der mCLOUD wurden in die Merkmalsmatrix eingetragen und diese darüber weiter präzisiert und erweitert. Die vorhandenen Informationen wurden analysiert und die Nutzbarkeit der Metadaten und Beschreibungsdaten evaluiert.

Die nachfolgende Darstellung zeigt die Benutzung der Merkmalsmatrix für die Analyse einer Datenpublikation der mCLOUD zu CO2-Emissionen, die von der Deutschen Bahn bereitgestellt wurde.

Anbieter			DB
Dimension	**Merkmal**	**Ausprägungen**	**CO2 Emissionen EiD_CO2_kg_WTW_Fpl2014_Grid2500_GK3**
Raum (Geometrie und Topologie)	Objekttyp	Raster Vektor (Punkt|Linie|Fläche|Volumen) Netzwerk	Polygon
	Objekttyp - Details		zur Abbildung von Grids
	räumliche Ausdehnung	global lokal (Region, Land, Stadt...)	Deutschland
Thematik	Domäne	Straßen Klima & Wetter Wasserstraßen & Gewässer Bahn Infrastruktur Luft- & Raumfahrt	Bahn
	Vokabular	kein Standard Standard	kein Standard
	Klassifikation	klassifiziert unklassifiziert	unklassifiziert
	Anzahl Themen	1 n	1
	Skalenniveau	nominal ordinal ratio intervall	ratio
	Erhebung / Berechnung	Einzelwert kumuliert gemittelt ...	
	Berechnung - Details		
Zeit	Zeitangabe (Erhebungszeit)	diskret Zeitpunkt diskret Zeitspanne kontinuierliche Zeitreihe	diskret Zeitpunkt
	Konkrete Zeitangabe		2014

	Zeitangabe enthalten in	Downloadseite Metadaten Dateiname Dateninhalt	Dateiname Downloadseite
	Genauigkeit (der Zeitangabe)	sekündlich minütlich stündlich täglich …	jährlich
	Auflösung (Periode/Intervall)	sekündlich minütlich stündlich täglich …	Jahr
	Gültigkeit	1-n Sekunden 1-m Minuten 1-o Stunden …	
	Ordnung	linear verzweigend zyklisch	
Meta	Format		Shape File, CSF
	Quelle der Daten	Portal (mCLOUD…)	
	Lizenz		
	Link zum Datensatz		
	weitere Informationen	(in Daten enthalten, findet aber in der Tabelle keine Abbildung)	

Tabelle 2: Beispielhafte Analyse eines Datensatzes in der Datenmatrix

Als Herausforderungen aus dieser Analyse resultierten:

- Das Fehlen der zeitlichen Merkmale in Metadaten stellt eine Herausforderung dar, so dass bereits hier deutlich wurde, dass zeitliche Informationen teilweise nur in den Daten selbst enthalten sind und die Analyse der Metadaten unbedingt ergänzt werden muss um die Analyse der Daten.
- Viele der Datenpublikationen sind unterschiedlich strukturiert, was besonders bei CSV-Dateien eine Algorithmen-gesteuerte Analyse erschwert, da hier mit sehr vielen Spezialfällen umgegangen werden muss.
- Eine sinnvolle Visualisierung ist teilweise stark auch von der Thematik wie z.B. Klima, Verkehr etc. abhängig, wofür sinnvolle Vokabulare nötig sind.
- Das Fokussieren auf die Zeitangabe alleine ist nicht ausreichend. Eine Ergänzung um die Erhebungszeit der Daten vs. der Metadaten ist unbedingt vorzunehmen.

- Zusätzliche Informationen über die Datenveredelung sind wichtig, wie die Unterscheidung nach Rohdaten, kumulierten Werten, Aggregationen etc.

2.3. Analyse raumzeitlicher Visualisierungen

Bei der Analyse raumzeitlicher Visualisierungen wurde eine Liste verschiedener Arbeiten erstellt, wie sie aktuell in wissenschaftlichen Publikationen, aber auch in der Datenvisualisierungs-Community verwendet werden.

Die erste Version der Liste umfasste 49 Arbeiten. Diese Visualisierungen wurden auf Grundlage ihrer Komplexität, Nutzerfreundlichkeit und Umsetzbarkeit in unserem Szenario sortiert und gefiltert so, dass wir letztendlich auf eine Liste von 29 verschiedenen Visualisierungen zurückgreifen können. Die vollständige Liste inklusive Abbildungen kann in Anhang A gefunden werden.

Diese Visualisierungen ließen sich in fünf Kategorien unterteilen, welche im Folgenden beschrieben werden.

2.3.1. 3D Glyphen

Diese Kategorie zeichnet sich durch die Kombination aus flachen 2D Karten in Verbindung mit darauf platzierten 3D Elementen aus. Dabei wird die räumliche Komponente durch die 2D Karte repräsentiert, während die zeitliche Komponente meist über die Höhe/z-Achse der 3D Glyphen abgebildet wird.

Mit Hilfe der Glyphen können dann eine Vielzahl an Variablen visualisiert werden. Ein Vorteil dieser Art der Visualisierung ist die zusätzliche Dimension, welche für die Darstellung verschiedener Thematischer Variablen genutzt werden kann. Einer der Nachteile ist allerdings auch, dass Elemente sich schnell überlagern und zum Beispiel Höhenunterschiede je nach Blickwinkel visuell schwerer zur unterscheiden sind.

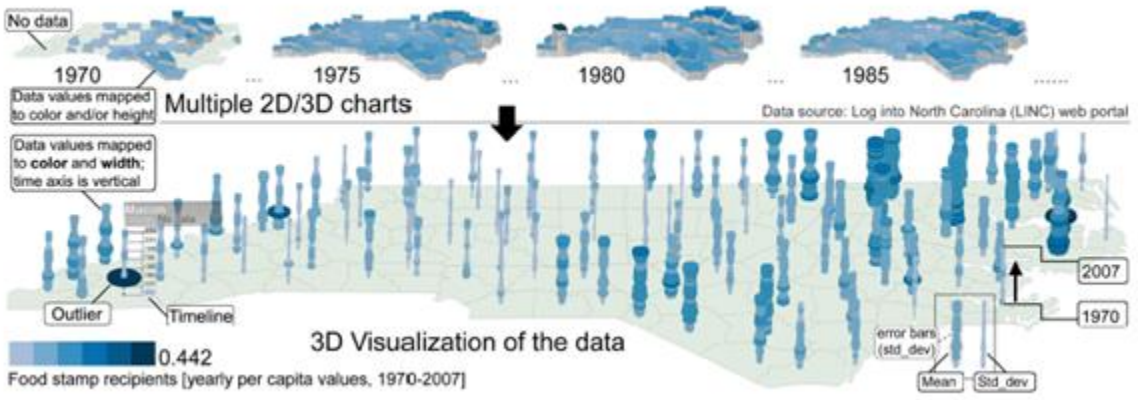


Abb. 8: Raumzeitliche Visualisierung der Verteilung von Essensmarken in North Carolina. Der obere Teil der Abbildung zeigt Small Multiples der verschiedenen Jahre, während der untere Teil 3D Glyphen auf einer 2D Karte zeigt, bei denen die zeitliche Achse vertikal abgebildet ist (Quelle siehe: Data Vases).

2.3.2. 3D Pfade

Ein ähnlich hybrides Konzept verfolgen Pfade- und Datenflüsse in 3D Visualisierungen. Auch hier dient eine 2D Karte als räumliche Basis auf welcher dann komplexere Visualisierungen aufbauen. Für die zeitliche Komponente wurden hier verschiedene Systeme gefunden. Mit einem Pfad als Grundstruktur wurden die zeitlichen Komponenten sowohl in Richtung der z-Achse als auch in Richtung der x-/y-Achsen abgetragen.

Mit der Zeit in x-/y-Richtung, bewegt sich die Thematische Variable entlang eines Pfades, d.h. es werden zum Beispiel Daten einer Autofahrt entlang dieses Pfades visualisiert. Bei den verschiedenen Variablen handelt es sich zum Beispiel um Geschwindigkeiten, Verkehrsaufkommen, oder Abgaswerte.

Alternativ finden sich in der Gruppe auch Visualisierungen, welche die zeitliche Komponente auf die z-Achse abbilden, so dass zum Beispiel ein Pfad kontinuierlich ansteigt und somit einen zeitlichen Ablauf aufzeigt.

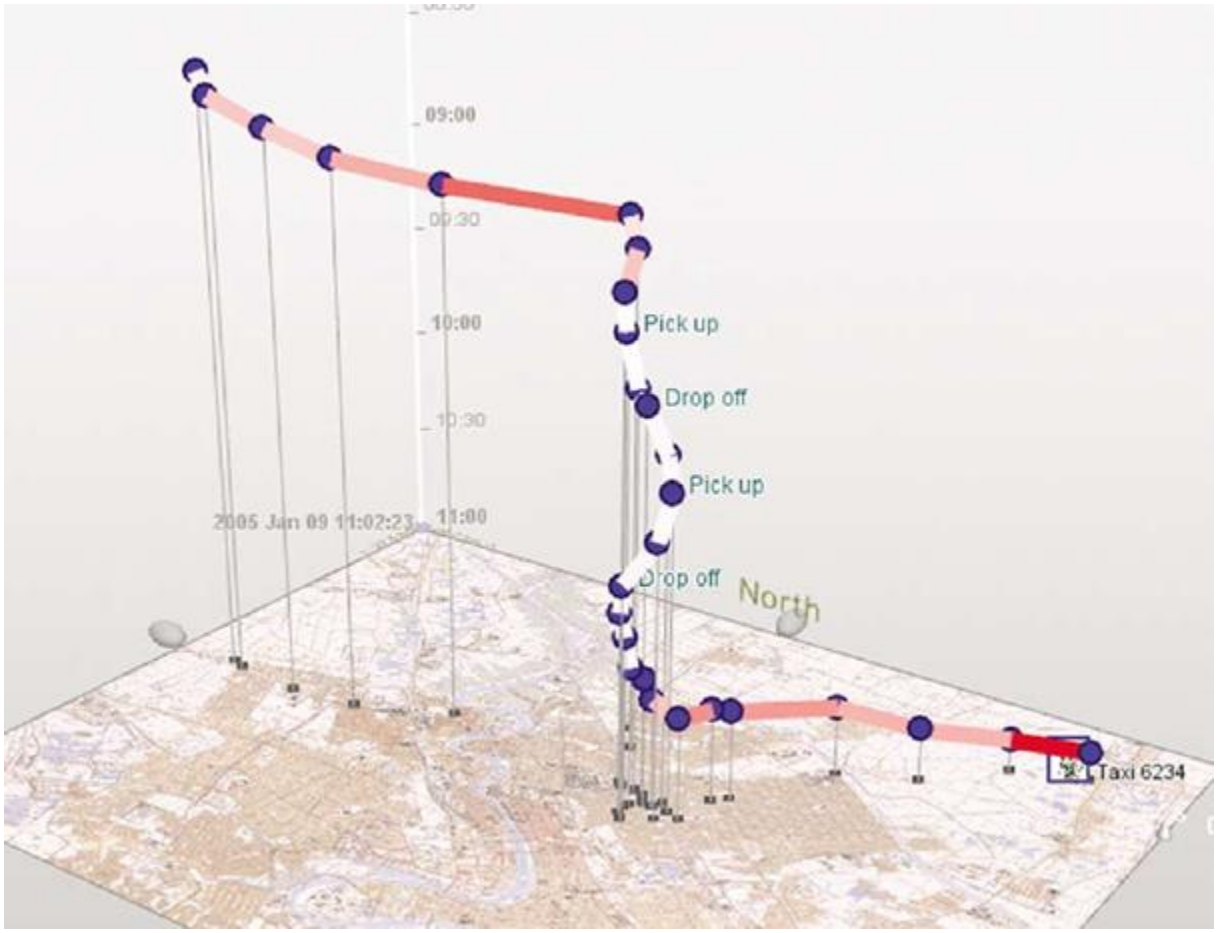


Abb. 9: Eine Visualisierung des Space-Time Cube Konzeptes. Die Linie gibt die Reihenfolge der Events vor, während die Höhe der Linie die Zeit visualisiert (Quelle siehe: GeoTime).

2.3.3. 2D Karten

Ein großes Feld raumzeitlicher Visualisierungen wird von 2D Karten beherrscht. Diese ermöglichen es komplexe Zusammenhänge auf den Karten selbst zu visualisieren, so dass räumliche und zeitliche Aspekte in derselben Ebene liegen. Durch die fehlende Dimension, werden zeitliche Aspekte sehr oft durch Animationen und Zeitleisten integriert, es gibt aber auch vereinzelte Methode, bei welchen die Zeitkomponente in einem Uhr-ähnlichen System angezeigt wird.

Ein Beispiel für diese Art von Visualisierung sind Flowmaps, welche von Verkehrsaufkommen über Bevölkerungsmigrationen und Flugzeugrouten die verschiedensten Daten visualisieren können.

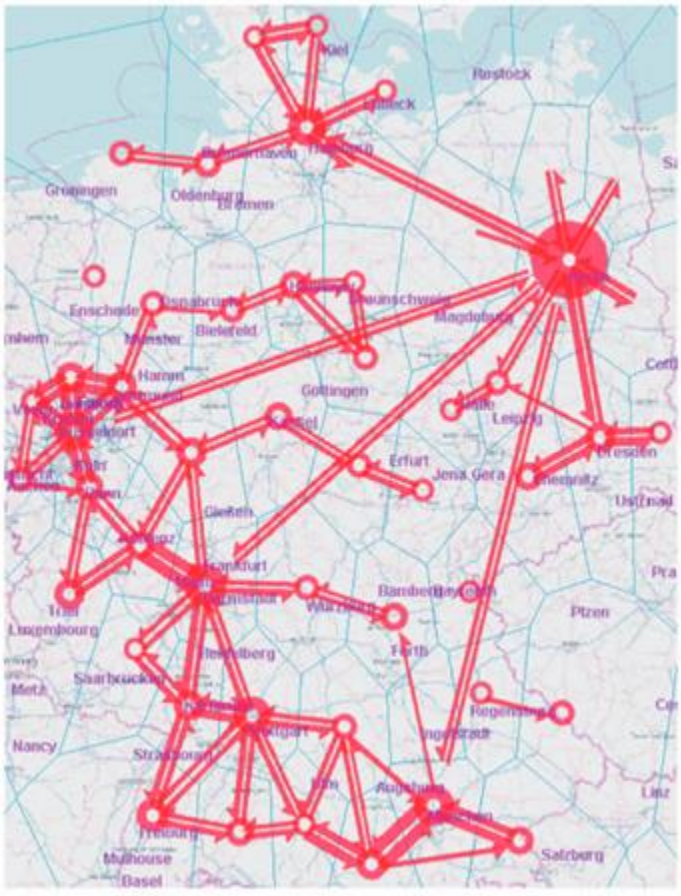


Abb. 10: Eine 2D Flow Map, die Richtung und Frequenz angibt (Quelle siehe Flow Map).

2.3.4. Kombination aus Karten und Diagrammen

Alternativ zu reinen Kartenvisualisierungen finden sich auch viele Kombinationen aus Karten und einfachen Diagrammen, wie zum Beispiel Balkendiagrammen und Scatterplots. Durch die Kombination mehrerer Visualisierungen eröffnen sich andere Möglichkeiten Dimensionalität in die Visualisierung zu integrieren.

Bei dieser Art der Visualisierung kann man in den meisten Fällen Punkte auf der Karte auswählen und erhält dazu verschiedene Daten, d.h. pro geographischem Punkt können verschiedene thematische Variablen angezeigt werden. Die räumlichen Aspekte werden bei dieser Visualisierungsmethode also hauptsächlich über die Karte abgedeckt, während zeitliche Aspekte in den einzelnen Diagrammen angegeben werden.

Ein Nachteil ist, dass diese Kombination ungeeignet für einen direkten geographischen Vergleich ist, solange man nicht mit Hilfe verschiedener Interaktionstechniken die Möglichkeit schafft mehrere Orte gleichzeitig zu betrachten. Ein Vorteil hingegen ist die große Vielfältigkeit im Hinblick auf zu visualisierende, thematische Variablen.

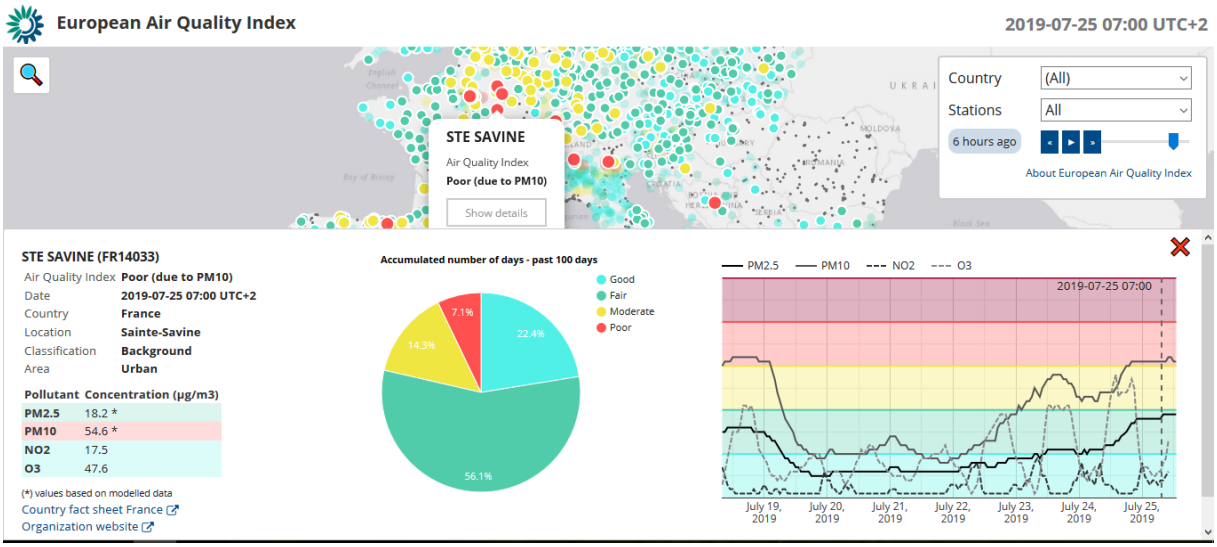


Abb. 11: Eine Kombination aus 2D Karte, Linien- und Kreisdiagramm zur Visualisierung der Luftqualität in Europa (Quelle: European Air Quality Index, European Environment Agency, 26.05.2020, http://airindex.eea.europa.eu/).

2.3.5. Sonderfälle

Neben den gängigen raumzeitlichen Visualisierungen konnten wir auch einzelne, interessante Sonderfälle finden. Bei diesen wurde der zeitliche Fokus stark in den Vordergrund gerückt, so dass sie weniger Möglichkeiten bieten räumliche Komponenten zu analysieren. Da diese Visualisierungen ohne klassische Karte arbeiten sind sie in unserer Analyse eine Randgruppe.

Einer der Sonderfälle ist zum Beispiel ein 3D Modell eines Gebäudes, welches einen sehr kleinen, lokalen Bezug bietet, im Großen und Ganzen jedoch mit Hilfe von Opazität und Farbgebung die zeitliche, bauliche Entwicklung visualisiert.

Andere Visualisierungen stellen zeitliche Ereignisse an verschiedenen Orten einander gegenüber und nutzen dabei Tile Maps oder Heatmaps mit einem einfachen, textuellen, räumlichen Label.

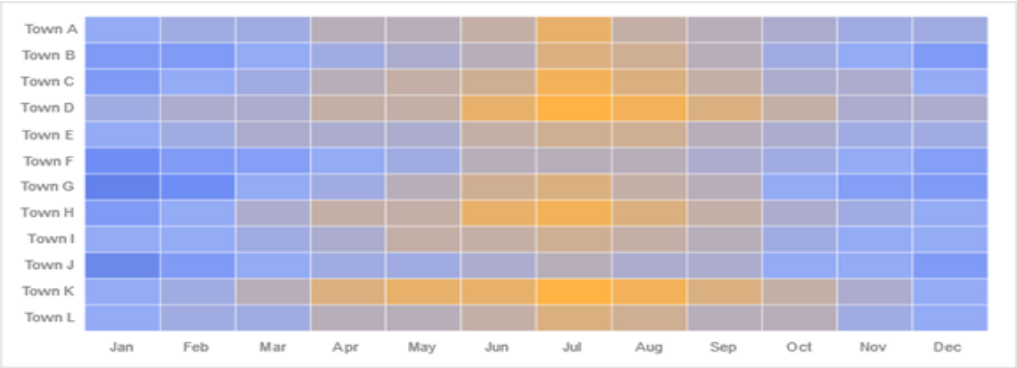


Abb. 12: Heatmap, die raumzeitliche Daten visualisiert ohne Verwendung einer Karte (Quelle siehe Heatmap (ohne Karte)).

3 Kategorien/Klassifizierung der Visualisierungen

Im nachfolgenden Kapitel wird beschrieben, wie die einzelnen Komponenten der Analyse in ein Gesamtkonzept integriert werden können.

3.1. Kombination Datenmatrix, Umfrage und Visualisierungsliste

Auf Grundlage der vorhergehenden Analyse war es möglich die einzelnen Komponenten zu klassifizieren und sowohl für die Daten der mCLOUD, als auch für die potentiellen, raumzeitlichen Visualisierungen übergreifende Muster zu entwickeln.

Ein essentieller Bestandteil unseres Konzeptes war es, aus diesen einzelnen Klassifizierungen ein Gesamtkonzept zu entwickeln und eine gemeinsame Sprache zu finden, welche über die fünf bereits vorgestellten Visualisierungskategorien, sowie die vier Datenkategorien hinausgeht.

Zu diesem Abgleich zwischen Datenmerkmalen und Visualisierungsaspekten haben wir drei primäre Fragen gestellt:

1. Welche Datenmerkmale benötigt man zur Wahl einer bestimmten mindestens Visualisierung?

   Die Frage zielt darauf ab, welche Merkmale ein Datensatz mindestens vorweisen muss, um eine bestimmte Visualisierung zu wählen. Welche Angaben sind also ausreichend um irgendeine Empfehlung zu geben.

2. Was sind zusätzliche Merkmale, um eine intelligente Auswahl zu gewährleisten?

   Da wir bei mVIZ Wert darauf legen ein Konzept zu entwickeln, welches den Entscheidungsprozess unterstützt und es Nutzern und Entwicklern ermöglicht eine Nutzerfreundliche und geeignete Visualisierungsauswahl zu treffen, sind zusätzliche Merkmale notwendig. Aus diesem Grund ist es entscheidend, welche ergänzenden Merkmale notwendig sind, um nicht nur eine bestimmte Visualisierung zu wählen, sondern innerhalb der Empfehlung sicherzustellen, dass Visualisierung und Datensatz sich gut ergänzen und zueinander passen.

   Diese Frage fokussiert sich darauf, welche Aspekte eine Visualisierung unter den gegebenen Voraussetzungen abbilden kann und wofür sie ungeeignet ist.

3. Was soll visualisiert werden?

   Abschließend geht es bei mVIZ aber nicht nur um die Eignung von Visualisierung und Datensatz, sondern auch um die Eignung von Visualisierung und Visualisierungsziel. Da die Wahl der Visualisierung essentiell darüber entscheidet, welche Frage beantwortet werden kann und welche nicht, ist es nötig auch die Nutzungsbedarfe direkt mit in die Entscheidung zu einer Auswahl einfließen zu lassen.

Durch Beantwortung dieser Fragen konnte eine Entscheidungsmatrix mit 12 spezifischen Kategorien und insgesamt 40 Ausprägungen erstellt werden. Zusammen bilden sie die Spalten unserer Visualisierungsmatrix und dienen als Entscheidungspunkte in unserem Empfehlungs-Prozess.

Die spezifischen Kategorien können in drei Meta-Kategorien unterteilt werden, welche im Folgenden im Detail vorgestellt werden.

3.2. Automatisch Generierte Entscheidungspunkte

Die Metakategorie „Automatisch Generierte Entscheidungspunkte“ basiert auf der Auswertung der verfügbaren Datensätze und der daraus resultierenden Klassifizierung, sowie einer Analyse bestehender Visualisierungen und deren Anforderungen.

Sie besteht aus sieben Kategorien mit 23 Ausprägungen, welche bestimmen was für Aspekte eines Datensatzes eine Visualisierung abbilden kann.

3.2.1. Anzahl thematischer Variablen

Die Visualisierung kann x unterschiedliche Variablen darstellen. Einige thematische Variablen wären zum Beispiel: Niederschlag, Lufttemperatur, Verkehrsaufkommen, oder die Geschwindigkeit eines Fahrzeugs.

Kann eine Visualisierung nur die räumlichen und zeitlichen Werte darstellen, wird der Datensatz in Time-Space-0 (TS-0) eingeordnet. Ist die Visualisierung einer einzelnen zusätzlichen thematischen Variable, wie Niederschlag, möglich, dann erfolgt eine Einordnung in TS-1 und TS-m, wenn mehrere zusätzliche, thematische Variablen visualisiert werden können.

Ausprägung	Definition
TS-0	Räumlich, Zeitlich
TS-1	Räumlich, Zeitlich, 1 zusätzliche Variable
TS-m	Räumlich, Zeitlich, X zusätzliche Variable

Tabelle 3: Definition der Anzahl thematischer Variablen

3.2.2. Zeitliche Überlagerung verschiedener Themen

Die Werte unterschiedlicher, thematischer Variablen zu x Zeitpunkten können innerhalb derselben Visualisierung dargestellt werden. Das bedeutet, die Visualisierung ist, je nach Ausprägung, in der Lage verschiedene Zeiträume darzustellen.

Für die Ausprägung *identisch* heißt das, dass der Zeitstempel der einzelnen Variablenmessungen annähernd identisch ist. Ein Beispiel wären mehrere Niederschlagsmessungen an verschiedenen Orten, welche zum immer gleichen Zeitpunkt (z.B.: 12:00 Uhr) stattfinden. Der abzubildende Zeitraum stimmt dementsprechend annähernd überein.

Ist der Zeitraum *überlappend* so hat ein Teil der Daten hat z.B. die Zeitstempel 01:00 Uhr und 12:00 Uhr, während der restliche Datensatz Werte für 12:00 Uhr und 15:00 Uhr bereit stellt.

Sind keine Übereinstimmungen bezüglich der zeitlichen Aspekte im Datensatz auszumachen, wird eine Visualisierung benötigt, die *getrennte* Zeiträume in einer Visualisierung darstellen kann.

<u>Abhängigkeiten innerhalb der Kategorien:</u>

- (*) Automatische Entscheidungspunkte à Anzahl thematische Variablen → TS-1 || TS-m

  Enthält der Datensatz lediglich raumzeitliche Aspekte (TS-0), ohne eine weitere thematische Variable (TS-1, TS-m), dann ist eine Unterscheidung der Überlappung, nicht notwendig.

- (optional) Nutzerinteraktion:
  - „Wie viele thematische Variablen sollen gleichzeitig visualisiert werden?“
  - „Wie viele Daten innerhalb einer Variable (z.B. unterschiedliche Trajektorien) sollen gleichzeitig visualisiert werden?“ [Nicht Teil der Matrix]

Ausprägung	Definition
Identisch	Zeitpunkte od. Zeitintervalle identisch
Überlappend	Zeitintervalle überlappen sich
Getrennt	Zeitpunkte od. Intervalle sind verschieden

Tabelle 4: Definition zeitlicher Überlagerung verschiedener Themen

3.2.3. Räumliche Überlagerung verschiedener Themen

Die Werte unterschiedlicher Variablen mit zugehöriger räumlicher Abdeckung können innerhalb derselben Visualisierung dargestellt werden. Diese Kategorie verhält sich gleich der vorhergehenden Ausprägung (Kapitel 3.2.2) mit Bezug auf die räumlichen Werte.

Abhängig von der räumlichen Ausdehnung (lokal, regional, global), die dargestellt werden soll und den darzustellenden räumlichen Einheiten (genaue Koordinate, Region, …), können auch hier Überlagerungen stattfinden.

<u>Abhängigkeiten innerhalb der Kategorien:</u>

- (*) Automatische Entscheidungspunkte à Anzahl thematische Variablen → TS-1 || TS-m

  Enthält der Datensatz lediglich raumzeitliche Aspekte (TS-0), ohne eine weitere thematische Variable (TS-1, TS-m), dann ist eine Unterscheidung der Überlappung, nicht notwendig.

- (optional) Nutzungskontext → räumliche Ausdehnung [Nicht Teil der Matrix]
  - Lokal (kleinräumig, Straßen, detailreich, max. Stadt)
  - Regional (mittelräumig, max. Land)
  - Global (weniger Kartendetails, max. Welt) → Überlappung *identisch* sehr wahrscheinlich
- (optional) Nutzerinteraktion:
  - „Wie viele thematische Variablen sollen gleichzeitig visualisiert werden?“
  - „Wie viele Daten innerhalb einer Variable (z.B. unterschiedliche Trajektorien) sollen gleichzeitig visualisiert werden?“ [Nicht Teil der Matrix]

Ausprägung	Definition
Identisch	räumliche Abdeckung identisch bzw. sehr ähnlich
Überlappend	räumliche Abdeckung überlappend
Getrennt	räumliche Abdeckung verschieden (unterschiedliche Orte, keine Art od. Anzahl von Attributen gleich am selben Ort)

Tabelle 5: Definition räumlicher Überlagerung verschiedener Themen

3.2.4. Skalenniveau

Die thematischen Variablen haben ein thematisches Skalenniveau, welches von der Visualisierung abgebildet werden kann. Während die Skalenniveaus in der Empirie hauptsächlich festlegen, welche mathematischen Operationen und Transformationen mit den Daten durchgeführt werden, zielt die Klassifizierung in diesem Fall hauptsächlich darauf ab, welche Skalenniveaus bei Verwendung einer speziellen Visualisierung noch dargestellt und abgelesen werden können.

Ausprägung	Definition
Nominal	qualitativ
Ordinal	qualitativ, Rangordnung, größer/kleiner etc.
Kardinal (Ratio/ Interval)	R: Quantitativ, metrisch, existiert ein absoluter Nullpunkt (20 Kelvin sind doppelt so heiß wie 10K) I: Quantitativ, metrisch, Rangunterschiede zwischen einzelnen Daten, kein absoluter Nullpunkt (Datum, Celsius…), z.b. 20C ist nicht doppelt so heiß wie 10 Grad C.

Tabelle 6: Definition Skalenniveau

3.2.5. Geometrietyp

Beim Geometrietyp handelt es sich um die Ausdehnung, auf die sich ein thematischer Datenpunkt beziehen (Geometrische Primitive entspr. ISO 19107).

Ausprägung	Definition
Punkt / Multipoint	Position, Koordinate, nicht ausgedehnter Ort, ggf. Aneinanderreihung von Positionen
Linie / Multiline	Strecke mit Start- und Endpunkt
Polygon / Multipolygon	Fläche
Volumen	3D Stadtmodelle
Trajektorie	Punktabfolge mit Zeitinformation
Coverages (Raster, TIN)	

Tabelle 7: Definition Geometrietyp

3.2.6. Fokus

Da nicht jede Visualisierung sowohl räumliche, als auch zeitliche Veränderungen anzeigen kann, sich viele jedoch mit Hilfe von Animationen o.ä. dazu anbieten, beschäftigt sich die Kategorie „Fokus“ damit, welche Veränderung von der Visualisierung standardmäßig dargestellt werden können.

<u>Abhängigkeiten innerhalb der Kategorien:</u>

- „Zeitabfolge” → ist dynamisch ausgewählt wird der Datensatz „raumzeitlich“ [Nicht Teil der Matrix]

  Hat ein Nutzer „dynamisch“ gewählt, so kann eine Visualisierung, welche nur entweder einen räumlichen oder zeitlichen Fokus bietet, im Zusammenhang auch einen raumzeitlichen Fokus bieten.

Ausprägung	Definition
Zeitlich	Zeitliche Veränderung der Thematik, keine räumliche Veränderung (z.B. fest installierte Temperaturmessstation)
Räumlich	Räumliche Veränderung der Thematik, keine zeitliche Veränderung (z.B. statische Phänomene, Momentaufnahme)
Raumzeitlich	Räumliche und zeitliche Veränderungen

Tabelle 8: Definition des Fokus der thematische Variablen

3.2.7. Zeitprimitiv

Welche zeitlichen Primitive werden von der Visualisierung unterstützt.

Ausprägung	Definition
Punkt (diskreter Punkt)	Punktmessung
Intervall (diskrete Spanne)	Zeitspanne / Intervall

Tabelle 9: Definition des zeitlichen Primitives

3.3. Nutzer-gestützte Entscheidungen

Da sich nicht alle Fragen automatisch durch den Datensatz beantworten lassen, es jedoch einer detaillierten Unterscheidung bedarf, um nutzerfreundliche Visualisierungen zu empfehlen, ist es notwendig einige Aspekte vom Nutzer entscheiden zu lassen.

Dafür wurden vier Kategorien mit sechs Ausprägungen im Konzept verankert, welche im Folgenden vorgestellt werden.

3.3.1. Zeitmodell

Lineare und zyklische Zeitdarstellungen können unterschiedliche Fragen beantworten und den Nutzer auf verschiedene Weisen unterstützen. Das Zeitmodell beschreibt dabei, wie die Zeit von der Visualisierung dargestellt werden kann. Es gibt zum einen die Option einer linearen, fortlaufenden Darstellung, wie man sie bei Liniendiagrammen findet, und zum anderen gibt es

die Option einer zyklischen Darstellung, wie sie z.B. bei Uhren oder in Spiral-Visualisierungen wiederfindet.

Mögliche Fortführung:

- (Optional) Nutzerabfrage bei „zyklisch“: „In welchem Zyklus sollen die Daten dargestellt werden“ → x Minuten/Monate/Jahre. [Nicht Teil der Matrix]

Ausprägung	Definition
Linear	Linear, gerade (fortlaufende) Abfolge der Daten
Zyklisch	Kreis, Spirale, Zeitliche Perioden sind ersichtlich (Jahreszeiten etc.)

Tabelle 10: Definition des Zeitmodells

3.3.2. Zeitabfolge

Die Zeitabfolge spielt sowohl bei der Auswahl der Visualisierungen, als auch bei den Präferenzen eines Nutzers eine starke Rolle. Ob Animationen erwünscht sind, kann ein Nutzer also in dieser Kategorie festlegen.

Ausprägung	Definition
Statisch	Statisch (Interaktionen zu statischen Inhalten möglich)
Dynamisch	Abbildung von Veränderungen z.B. durch Animation oder Interaktionen und Effekte.

Tabelle 11: Definition der Zeitabfolge

3.3.3. Dimension der Visualisierung

Wie in Kapitel 2.4 gezeigt, haben 2D und 3D Visualisierungen ihre Vor- und Nachteile. Aus diesem Grund ist es wichtig einem Nutzer die Entscheidung zu überlassen.

Ausprägung	Definition
2D	Visualisierung innerhalb einer Ebene
3D	Visualisierung innerhalb eines Raumes

Tabelle 12: Definition der Dimensionalität der Visualisierung

3.3.4. Anzahl der zu visualisierenden Variablen

Entscheidend für die Visualisierung ist es zudem, wie viele verschiedene thematische Variablen gleichzeitig visualisiert werden sollen. Dies kann jedoch nicht automatisch aus den Daten herausgelesen werden, da die Automatik immer von den maximal vorhandenen Daten ausgeht. [Nicht Teil der Matrix, sollte jedoch abgefragt werden]

<u>Abhängigkeiten innerhalb der Kategorien:</u>

- Anzahl Variablen → TS-m (sonst ergibt eine Auswahl keinen Sinn)

Ausprägung	Definition
1	Platzhalter – Abhängig vom Datensatz. Auswahlliste inkl. Themen wäre hier sinnvoll.
m	Platzhalter – Abhängig vom Datensatz. Auswahlliste inkl. Themen wäre hier sinnvoll.

Tabelle 13: Anzahl der zu visualisierenden Variablen

3.4. Empfehlungskategorien

3.4.1. Intentionsunterstützung

Eine Visualisierung unterstützt das Erkennen bestimmter Aspekte eines Datensatzes und nicht jede Visualisierung ist dabei für jeden Aspekt geeignet. So ist eine Visualisierung, die *Muster* innerhalb eines Datensatzes erkennbar macht, meist nicht dazu geeignet *Verhältnisse* innerhalb der Daten aufzuzeigen.

Aus diesem Grund enthält die Visualisierungsmatrix einen Abschnitt zur Klassifizierung von Visualisierungen mit Hinblick auf mögliche Nutzerintentionen. Diese können entweder über eine Abfrage in die Empfehlungen mit einfließen, oder als zusätzliche Informationen in der Ergebnisübersicht angeboten werden.

Ausprägung	Definition
Verhältnisse (Proportions)	Unterschiede/Ähnlichkeiten anhand von Größe oder Fläche zwischen Werten oder *Teile des Ganzen*
Vergleiche (Comparisons)	Aufzeigen von Wertunterschieden oder -ähnlichkeiten
Verteilungen (Distribution)	Aufzeigen von Häufigkeiten, wie Daten verteilt oder gruppiert sind.
Muster	Aufzeigen von Formen und Mustern innerhalb der Daten
Beziehungen	Beziehungen und Korrelationen zwischen zwei oder mehr Variablen
Entwicklung über Zeit	Aufzeigen von Trends und Entwicklungen über Zeit
Flow	Informationsfluss, Bewegungen von einem Ort zum anderen, z.B. Verkehrsflüsse, Migration von Menschen, Tieren oder Gütern

Tabelle 14: Empfehlungskategorien basierend auf Nutzerintentionen

4 Praktische Umsetzung

4.1. Vorstellung der Matrix anhand eines Beispiels

Im Folgenden wird anhand eines Beispiels aus der mCLOUD die Verwendung der Visualisierungsmatrix demonstriert. Die Auswahl des Datensatzes erfolgte zufällig, jedoch unter der Bedingung, dass er raumzeitliche Daten enthält.

4.1.1. Analyse des Datensatzes

Der Datensatz wurde dann mit Hilfe der Entscheidungspunkte händisch ausgewertet. Dieser Schritt wird im fertigen Demonstrator halbautomatisch ausgeführt, das heißt vorhandene Ausprägungen werden automatisch ermittelt und ausgewählt. Da jedoch bei aktuellen Datensätzen keine Standardisierung der Angaben vorherrscht, ist ein vollständiges, automatisches ausfüllen der Visualisierungsmatrix zu diesem Zeitpunkt unwahrscheinlich.

Bei dem ausgewählten Datensatz handelt es sich um „Historische 10-minütige Stationsmessungen der Max/Min-Temperaturen in 5cm und 2m Höhe in Deutschland“ [Link].

Die Auswertung erfolgt über die Entscheidungspunkte der Kategorie “Automatisch Generierte Entscheidungspunkte“.

Kategorie	Ausprägung	Datensatz-klassifikation	Begründung
#thematisch Variable	TS-0		• Temperaturmessung in 5cm Höhe • Temperaturmessung in 2m Höhe
	TS-1		
	TS-m	x	
Zeitliche Überlagerung der Themen	identisch	x	• Messungen fanden zu annähernd gleichen Zeitpunkten statt
	überlappend		
	getrennt		

Kategorie	Ausprägung	Datensatz-klassifikation	Begründung
Räumliche Überlagerung der Themen	identisch	x	• Messungen fanden an gleichen Stationen statt
	überlappend		
	getrennt		
Skalenniveau	Nominal		• Temperatur ist rational skalierbar
	Ordinal		
	Ratio/Interval	x	
Geometrietyp	Punkt	x	• Messung bezieht sich auf einen Punkt
	Linie		
	Polygon		
	Volumen		
	Trajektorie (Punktabfolge)		
	Coverages (Raster, TIN)		
Fokus	Zeitlich	x	• es findet eine zeitliche Veränderung der thematischen Variablen statt, während die Stationen gleich bleiben
	Räumlich		
	Raumzeitlich		

Kategorie	Ausprägung	Datensatz-klassifikation	Begründung
Zeitprimitiv	Punkt		• bei den Messungen handelt es sich um 10 Minuten Intervalle
	Intervall	x	

Tabelle 15: Beispielhafte Analyse eines Datensatzes mit der Visualisierungsmatrix

4.1.2. Abgleich der Analyseergebnisse mit Hilfe der Matrix

Mit Hilfe der aus dem Datensatz generierten Entscheidungspunkte, jedoch ohne Nutzereingaben, kann nun ein erster Abgleich mit der Visualisierungsmatrix durchgeführt werden.

Dabei entsteht die folgende Liste mit geeigneten Visualisierungen:

Kategorie	**Visualisierung**
3D Icons on maps	• Pencil Icons on Maps • Helix Icons on Maps • Wakame (3D Radar Chart)
Kombination mit Karten	• Heatmap (plus Karte) • Stacked Barchart • Stacked Area Chart

Tabelle 16: Ergebnisübersicht der Datenanalyse

Diese Empfehlungen können dann anhand verschiedener Nutzerbedürfnisse (z.B. 2D oder 3D) und Visualisierungsintention weiter eingegrenzt werden.

4.2. Sammlung aktueller Bibliotheken /Datenformate

Um eine Umsetzung von ausgewählten 3D-Visualisierungen für den Demonstrator realisieren zu können, wurden aus dem umfangreichen Angebot der zur Verfügung stehenden Techniken und Bibliotheken die nachstehenden genauer untersucht:

- die vektororientierte Bibliothek D3,
- die kartenorientierte Bibliothek Leaflet,
- die Javascript-Bibliothek Mappa,
- die Spieleengine Unity mit Mapbox-Plugin,
- die Anwendung Kepler sowie
- das Frontend Turfjs.

D3 (https://d3js.org/)

D3 ist eine beliebte vektororientierte Bibliothek, welche sich sehr gut für die Verarbeitung von 2D Daten anbietet. Für 3D-Daten ist D3 nur begrenzt zu empfehlen. Erschwerend hinzu kommt, dass D3 keinen nativen Support für Geodaten aufweist.

Beispiel: https://observablehq.com/@mbostock/floating-landmasses

```
canvas = {
  const context = DOM.context2d(width, height);

  const projection = d3.geoOrthographic()
      .translate([width / 2, height / 2])
      .precision(0.5);

  const path = d3.geoPath(projection, context);

  while (1) {
    context.clearRect(0, 0, width, height);
    projection.rotate([Date.now() * -2e-3, -15]);

    projection.scale(width / 2.3).clipAngle(90);
    context.beginPath();
    path(sphere);
    context.lineWidth = 1.5;
    context.strokeStyle = "#000";
    context.stroke();
    context.beginPath();
    path(land);
    context.filter = "blur(6px)";
    context.fillStyle = "rgba(0,0,0,0.4)";
    context.fill();
    context.filter = "none";
    context.beginPath();
    path(graticule);
    context.lineWidth = 0.5;
    context.strokeStyle = "rgba(0,0,0,0.2)";
    context.stroke();

    projection.scale(width / 2.2).clipAngle(107);
    context.beginPath();
    path(land);
    context.fillStyle = "#737368";
    context.fill();

    projection.scale(width / 2.2).clipAngle(90);
    context.beginPath();
    path(land);
    context.fillStyle = "#dadac4";
    context.fill();

    yield context.canvas;
  }
}
```

Abb. 13: Das obige Code Beispiel rechts erzeugt die Ausgabe links.

Leaflet (https://leafletjs.com/)

Leaflet ist eine kartenorientierte Bibliothek für HTML5, welche auf dem Open Street Map-Projekt aufsetzt. Die Bibliothek ist nur für 2D geeignet.

Beispiel: https://leafletjs.com/examples/choropleth/example.html

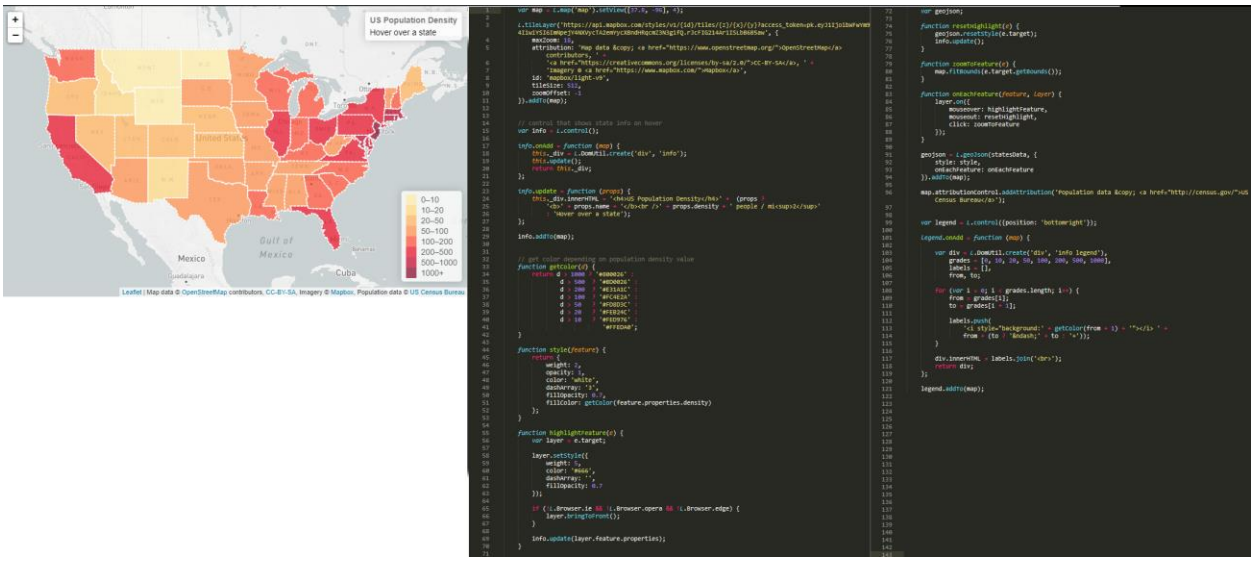
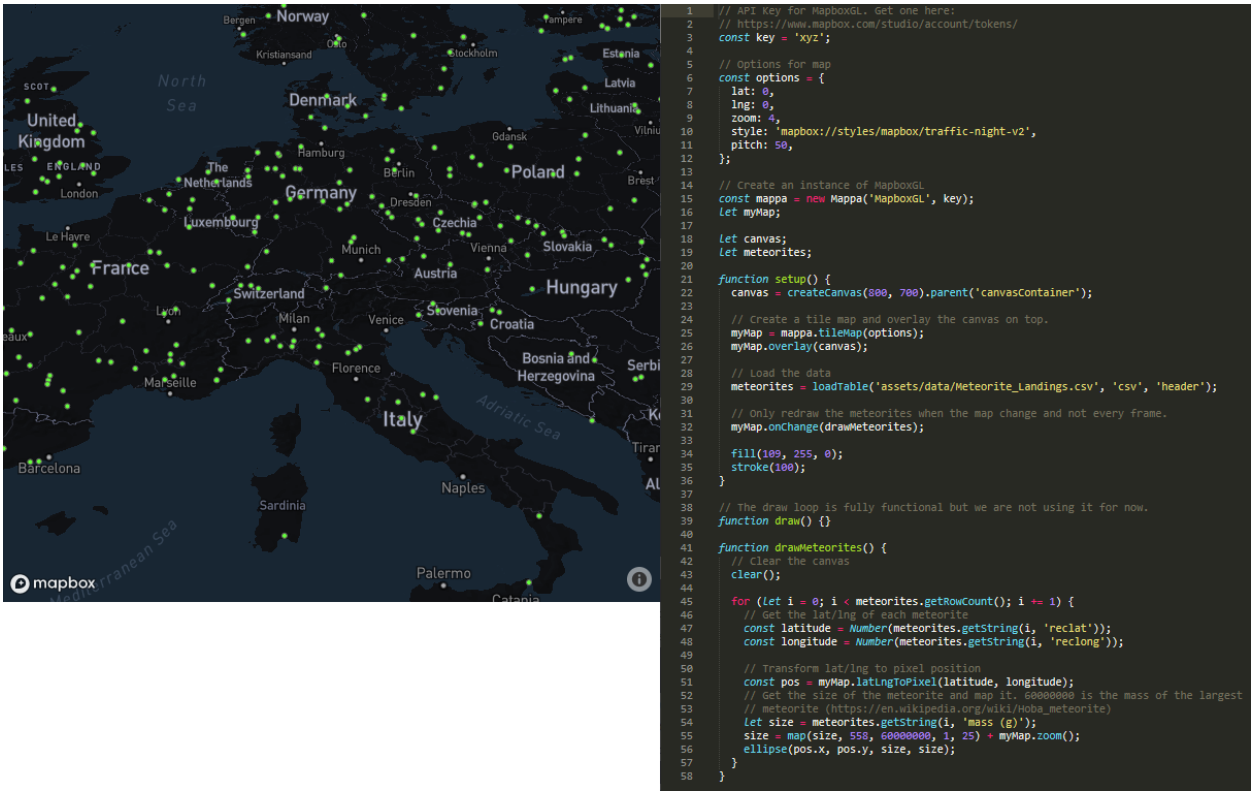


Abb. 14: Das obige Code Beispiel rechts erzeugt die Ausgabe links

Mappa (https://mappa.js.org/)

Mappa ist eine Javascript Bibliothek, die sowohl Google Maps als auch Mapbox als Karten-Backend nutzen kann.

Beispiel: https://mappa.js.org/docs/examples-mapboxGL.html




```
// API Key for MapboxGL. Get one here:
// https://www.mapbox.com/studio/account/tokens/
const key = 'xyz';

// Options for map
const options = {
  lat: 0,
  lng: 0,
  zoom: 4,
  style: 'mapbox://styles/mapbox/traffic-night-v2',
  pitch: 50,
};

// Create an instance of MapboxGL
const mappa = new Mappa('MapboxGL', key);
let myMap;

let canvas;
let meteorites;

function setup() {
  canvas = createCanvas(800, 700).parent('canvasContainer');

  // Create a tile map and overlay the canvas on top.
  myMap = mappa.tileMap(options);
  myMap.overlay(canvas);

  // Load the data
  meteorites = loadTable('assets/data/Meteorite_Landings.csv', 'csv', 'header');

  // Only redraw the meteorites when the map change and not every frame.
  myMap.onChange(drawMeteorites);

  fill(109, 255, 0);
  stroke(100);
}

// The draw loop is fully functional but we are not using it for now.
function draw() {}

function drawMeteorites() {
  // Clear the canvas
  clear();

  for (let i = 0; i < meteorites.getRowCount(); i += 1) {
    // Get the lat/lng of each meteorite
    const latitude = Number(meteorites.getString(i, 'reclat'));
    const longitude = Number(meteorites.getString(i, 'reclong'));

    // Transform lat/lng to pixel position
    const pos = myMap.latLngToPixel(latitude, longitude);
    // Get the size of the meteorite and map it. 60000000 is the mass of the largest
    // meteorite (https://en.wikipedia.org/wiki/Hoba_meteorite)
    let size = meteorites.getString(i, 'mass (g)');
    size = map(size, 558, 60000000, 1, 25) + myMap.zoom();
    ellipse(pos.x, pos.y, size, size);
  }
}
```

Abb. 15: Das obige Code Beispiel rechts erzeugt die Ausgabe links.

Unity mit Mapbox SDK (https://unity.com/ https://docs.mapbox.com/unity/)

Unity ist eine Spiele-Engine, die mit einem Mapbox-Plugin verfügbar und nutzbar ist. Bei Unity handelt es sich nicht um eine Javascript Bibliothek, sondern um eine weit verbreitete Spiele-Engine, in der man Anwendungen in C# programmieren kann. Unity ist eine der Spiele-Engines, welche sich in den letzten Jahren zu einem Standardwerkzeug im Bereich 3D Visualisierung von interaktiven Daten und Computerspielen etabliert hat. Unity hat eine sehr aktive Online-Community und bietet eine Vielzahl an Code-Bibliotheken. Aufgrund dieser Charakteristika wurde im Projekt für die Verarbeitung von 3D-Daten auf dieses Tool gesetzt. Unity bietet zudem ein natives Plugin für Mapbox, was als Kartenbackend genutzt werden kann. Unity übersetzt C#-Anwendungen nach HTML5. Die daraus entstehenden HTML5- Anwendungen sind übersichtlich groß und mit den meisten modernen Browsern Kompatibel. Leider ist das Deployment von Unity Anwendungen zum Teil kompliziert, weil zum Erstellen der HTML5-Anwendungen ein separates Tool benötigt wird.

Beispiel: https://docs.mapbox.com/unity/maps/overview/location-based-games/

Abb. 16: Zahlreiche Interaktive Anwendungen wie das obige “Astronauten Spiel” sind als Pakete im SDK mit eingebunden.

Kepler (https://kepler.gl/ & https://deck.gl/)

Kepler ist eine in sich abgeschlossene Softwarelösung die wiederum auf deck.gl und Mapbox aufsetzt. Kepler bietet viele in sich abgeschlossene Module für 3D-Karten und GeoJson Unterstützung. Leider ist die Dokumentation der Anwendung zum aktuellen Zeitpunkt unvollständig, weswegen es sich als sehr schwierig herausstellte, Kepler mit den anderen Bibliotheken und Anwendungen zu vergleichen.

Turfjs (http://turfjs.org/)

Turfjs ist ein reines visuelles Frontend für kartenbasierte Polygon-Formen und Linien-Typen. Es handelt sich hierbei um keine abgeschlossene Lösung, sondern um ein Zusatzmodul für vorhandene HTML5 Anwendungen. Turfjs nutzt Mapbox als Karten-Backend.

4.3. Implementierungen

4.3.1. Metadatenextrations- und Datenexplorationstool

Um die Datensätze der mCLOUD zu analysieren, wurde ein Metadatenextraktions- und Datenexplorationstool (MDET) entwickelt. In der folgenden Abbildung sind die einzelnen Komponenten der Architektur dieses Tools zu sehen.

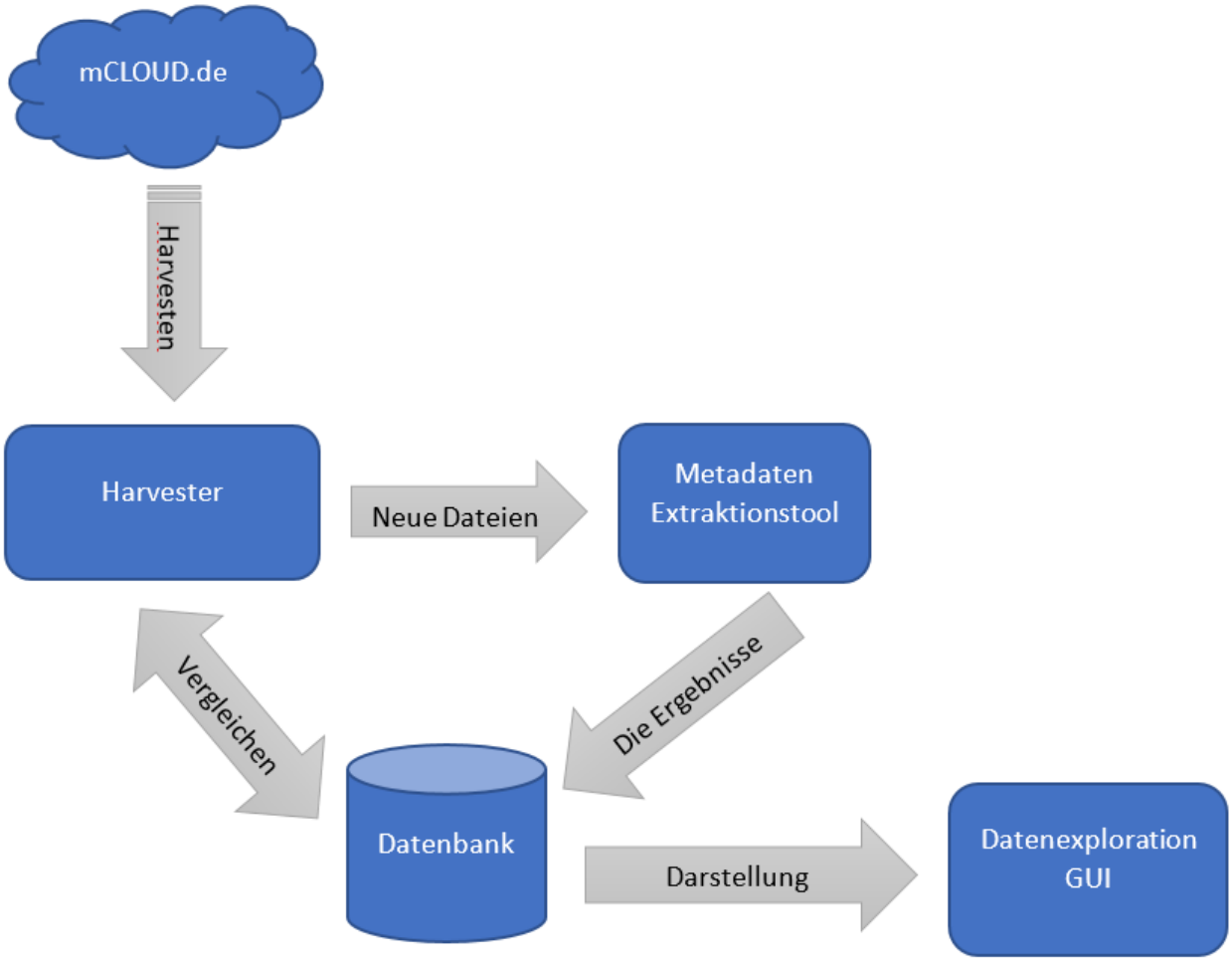


Abb. 17: Architekturkomponenten des Metadatenextraktionstools MDET

Wie in der Abb. 17 dargestellt, besteht dieses Tool aus vier Komponenten, die nachfolgend einzeln vorgestellt werden sollen.

Datenbank

Die CSV Dateien, die sich auf der mCLOUD befinden, sind statische Daten. Demzufolge ist es unnötig, sie in Echtzeit zu analysieren. Performanter und effizienter ist es, sie einmal zu analysieren (idealerweise gleich beim Hinzufügen) und die Ergebnisse der Analyse in einer Datenbank zu speichern.

Dafür wurde eine MongoDB Datenbank eingerichtet, die aktuell auf dem Cloud Dienst der MongoDB gehostet wird. In dieser Datenbank gibt es drei Kollektionen. Nachfolgend werden die Kollektionen, deren Zweck und das jeweilige Datenschema vorgestellt.

- Kollektion Datensätze:

Diese Kollektion ist ein Klon der mCLOUD Datenbank (mit genau denjenigen Informationen, die für die Analyse relevant und interessant sind) und dient dazu, alle Datensätze im Blick zu haben und die Veröffentlichung neuer Datensätze zu registrieren. Dadurch können die für neue Datensätze nötigen Analyse- und Verarbeitungsschritte direkt bei Verfügbarkeit eines neuen Datensatzes in der mCLOUD vorgenommen werden. Neben einer eigenen ID zur Identifizierung werden die ID aus der mCLOUD und deren Name sowie Metadaten wie Veröffentlichungsdatum, Anbieter und Dateiformat abgelegt. Nachfolgend ist der Aufbau dieser Kollektion skizziert:

```
{
    _id: {    type: ObjectId, required: true },
    cloudID:          {type: String,    required: true},
    cloudLink:        {type: String,    required: true },
    datasetName: {type: String,    required: true },
    issueTime:        {type: String },
    modificationTime:{type: String },
    publisher:        {type: String },
    isOnline:         {type: Boolean },
    distributions:[
        {
        distributionLink: {type: String,          required: true},
        distributionID:           {type: ObjectId,          required: true },
        fileFormat:       {type: String },
        downloaded:   {type: Boolean },
        analysisStatus: { type: String,  enum: [“pending”,"failed", "succeeded"]  }
        }
    ]
}
```

- Kollektion Exploration:

Diese Kollektion beinhaltet die Ergebnisse, die durch das MDET geliefert werden. Sie dient dazu, Inhalte jeder einzelnen CSV Datei explorieren zu können und stellt diese Informationen im Demonstrator über das User Interface bereit. Neben der Bezugnahme auf die ID aus der Kollektion Datensätze werden u.a. Informationen zu raumzeitlichen Daten der Analyse in dieser Kollektion abgelegt. Nachfolgend wird der Aufbau dieser Kollektion skizziert:

```
{
    _id:            {type: Objectid,    required: true    },
    cloudID:        {type: String,      required: true    },
    downloadLink:   {type: String,      required: true    },
    name:           {type: String,      required: true    },
    general:        {type: Object    },
    columns:        [{type: Object  }],
    overal_spatial: {type: Object    },
    overal_temporal: {type: Object  },
}
```

- Kollektion DCTMetadata:

Diese Kollektion beinhaltet die Metadaten, die von der Entscheidungsmatrix benutzt werden. Sie dient dazu, für jede einzelne CSV Datei passende Visualisierungsformen empfehlen zu können. Insofern werden hier die Analysedaten der Metadatenanalyse, die für die Identifizierung der Visualisierungsform relevant sind, abgelegt. Dazu gehören z.B. Aussagen zur zeitlichen Überschneidung, zum Skalenniveau oder auch der Anzahl an abgebildeten Themen. Der Aufbau dieser Kollektion ist nachfolgend skizziert:

```
{
    name:           { type: String, required: true },
    cloudID:        { type: String, required: true },
    fileID:         { type: String, required: true },
    thematicVariables: [{ type: String, enum: ["TS-0", "TS-1", "TS-m"] }],
    temporalOverlay: [{ type: String, enum: ["identical", "overlapping", "separated"] }],
    spatialOverlay: [{ type: String, enum: ["identical", "overlapping", "separated"] }],
    scaleOfMeasure: [{ type: String, enum: ["nominal", "ordinal", "ratio"] }],
```

geometryType: [{ type: String, enum: ["point", "line", "polygon", "volume", "trajectory", "coverage"] }],

focus: [{ type: String, enum: ["temporal", "spatial", "spatioTemporal"] }],

timePrimitive: [{ type: String, enum: ["point", "interval"] }]

}

Harvester

Der Harvester ist diejenige Komponente im Metadaten-Extraktions-Tool, die den Zugriff auf die mCLOUD realisiert. Ähnlich einem Web-Crawler werden die Datensätze der mCLOUD angefragt und mit der vorhandenen Datenbank verglichen. Werden neue Datensätze gefunden, werden diese in die Datenbank eingetragen und zur Analyse an das Extraktionstool übergeben. Das Extraktionstool analysiert jeden neuen Datensatz, wie nachfolgend beschrieben und trägt die Analyseergebnisse in die Datenbank ein.

Metadaten-Extraktions-Tool

Eine der Kernkomponenten ist das Metadaten-Extraktions-Tool. Dieses Tool ist in Python geschrieben und ermöglicht es, aus jeder CSV-Datei automatisch einheitliche, inhaltsbezogene Metadaten zu generieren. Darüber hinaus liefert es Informationen über die raumzeitlichen Aspekte (wenn vorhanden) jeder Datei.

Zurzeit ist das Tool limitiert auf CSV-Dateien, künftig kann es für andere Datentypen angepasst und somit erweitert werden. Es bekommt eine CSV-Datei als Input und generiert daraus Metadaten im JSON-Format.

Das Metadaten-Extraktionstool realisiert die Analyse in den nachfolgend angegebenen Schritten:

- Generelle Analyse:
    - Mit der Python-Funktion sniff: sample wird zunächst eine Objekt-Repräsentation der CSV-Datei erstellt. Sie enthält u.a. den Dialekt, das Encoding und die Überschriften.
    - Die Anzahl der Zeilen und Spalten wird analysiert.
- Prüfung der Datenqualität und Umsetzung der Datenbereinigung:
    - Identifizierung unerlaubter Zeichen gemäß dem Zeichensatz (durch fehlerhafte Enkodierung von deutschen Buchstaben),
    - Identifizierung von sprachlichen Unstimmigkeiten (z.B. verschiedene Formate bei Zahlen oder Datumsangaben),
    - Identifizierung von numerischen Angaben (Zahlen in Deutsch / Englisch),
    - Auffinden von leeren Spalten,
    - Identifizierung des Vorhandenseins unerwünschter Texte (am Anfang / Ende).
- Statistische Analyse jeder Spalte anhand der folgenden Kriterien bzw. Operatoren:

    - Min, Max, Mittelwert, Median, Q1 etc.,
    - Fehlende Werte,
    - Eindeutigkeit von Werten,
    - Häufigkeit der Werte,
    - Statistische Verteilung,
    - Bivariate Korrelation und Kovarianz.
- Räumliche Analyse:
    - Vorhandensein von räumlichen Daten,
    - Art der räumlichen Angaben,
    - Räumliche Ausdehnung.
- Zeitliche Analyse:
    - Vorhandensein von zeitlichen Daten,
    - Art der zeitlichen Angaben,
    - Zeitliche Ausdehnung.

Datenexplorationstool - Nutzerschnittstelle

Wenn ein Endbenutzer der mCLOUD sich einen Überblick über die Inhalte einer Datei wünscht, hat er bisher drei Informationsquellen zur Verfügung: die RDF-Metadaten auf der mCLOUD, die zugehörige Datensatzbeschreibung, und letztendlich die Datei. Aus diesen verschiedenen Informationsquellen resultieren entsprechende Probleme, wie nachfolgend dargestellt.

- RDF-Metadaten:
    - Es ist kein inhaltlicher Bezug gegeben.
    - sie sind für die maschinelle Verarbeitung entwickelt und somit schwer von Menschen lesbar.
    - Sie sind unvollständig.
    - Sie sind uneinheitlich.
    - Sie sind oftmals fehlerhaft (von Menschen eingegeben).
    - Sie sind schwer generierbar.
- Datensatzbeschreibung:
    - Die Datensatzbeschreibung ist lang, unverständlich und teilweise verwirrend.
    - Die Datensatzbeschreibung ist nicht vorhanden.
    - Die Datensatzbeschreibung ist unvollständig.
    - Die Datensatzbeschreibung ist uneinheitlich.
    - Die Datensatzbeschreibung ist (maschinell) schwer analysierbar.
- Die Datei selbst:
    - Die Datei muss für die weitere Verarbeitung zunächst heruntergeladen werden. Besonders problematisch stellt sich dies bei großen Datensätzen und bei Verwendung von Mobilgeräten dar.

- Um die Dateiinhalte zu analysieren ist eine manuelle Verarbeitung und Analyse erforderlich. Je nach Datei und -inhalt kann dies sehr mühsam werden. Besonders mit einem allgemeinen Datentyp wie in unserem Fall CSV, sind die Variationen nahezu unzählig, vor allem für raumzeitliche Daten. Zudem ist für jedes einzelne Format ein adäquates Tool notwendig.

Um diesen Prozess für die Benutzer zu erleichtern, werden die durch das Metadatenextraktionstool generierten Metadaten mittels einer grafischen Benutzeroberfläche visuell dargestellt. Diese Oberfläche wird in Form einer React-Webanwendung (Javascript) zur Verfügung gestellt.

Die dem Benutzer angebotenen Visualisierungen der Metadaten umfassen sowohl triviale Informationen (wie Format, Größe etc.) als auch spezifische Informationen, die für die raumzeitliche Analyse erforderlich sind. Die folgenden Abbildungen stellen die Nutzerschnittstelle auszugsweise dar. Abbildung 18 zeigt zunächst die allgemeinen Metadaten des gewählten Datensatzes wie Dateigröße, Dateiformat, Anzahl an Zeilen, Anzahl an Spalten. Zusätzlich wird ein Explorieren anhand eines Samples des Datensatzes ermöglicht.

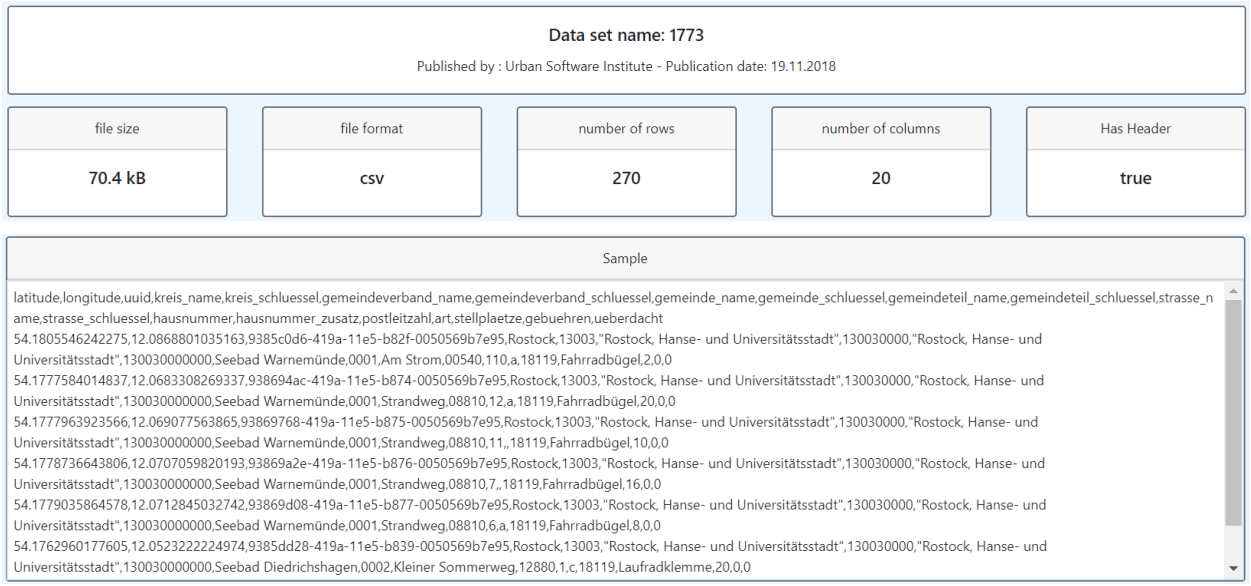
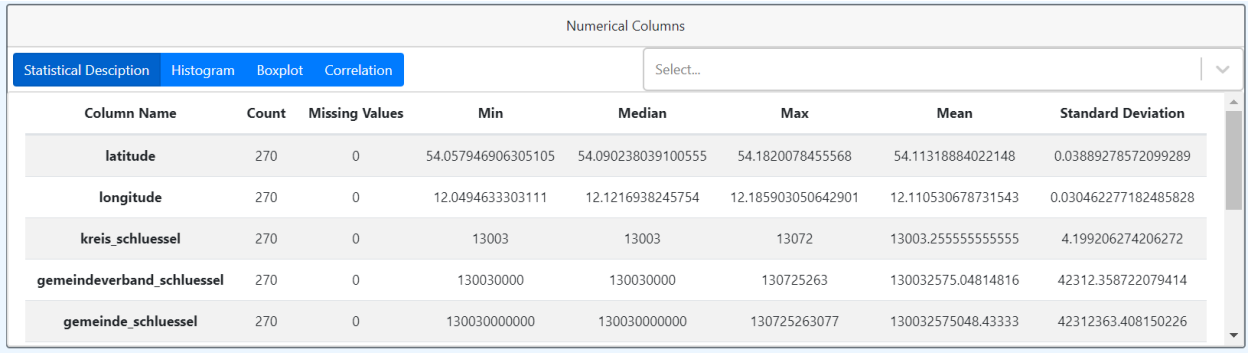


Abb. 18: Darstellung allgemeiner Metadaten zum gewählten Datensatz durch die Nutzerschnittstelle

Die Abbildungen 19 und 20 beziehen sich auf die Auswertung von numerischen Werten im ausgewählten Datensatz und zeigen einerseits einige statistische Daten wie Anzahl der Spaltenwerte, vorhandene Datenlücken, Minimum, Maximum, Median, Standardabweichung etc.



Column Name	Count	Missing Values	Min	Median	Max	Mean	Standard Deviation
latitude	270	0	54.057946906305105	54.090238039100555	54.1820078455568	54.11318884022148	0.0388927857209928
longitude	270	0	12.0494633303111	12.1216938245754	12.185903050642901	12.110530678731543	0.030462277182485828
kreis_schluessel	270	0	13003	13003	13072	13003.255555555555	4.199206274206272
gemeindeverband_schluessel	270	0	130030000	130030000	130725263	130032575.04814816	42312.358722079414
gemeinde_schluessel	270	0	130030000000	130030000000	130725263077	130032575048.43333	42312363.408150226

Abb. 19: Darstellung statistischer Daten zu numerischen Spalten des gewählten Datensatzes durch die Nutzerschnittstelle

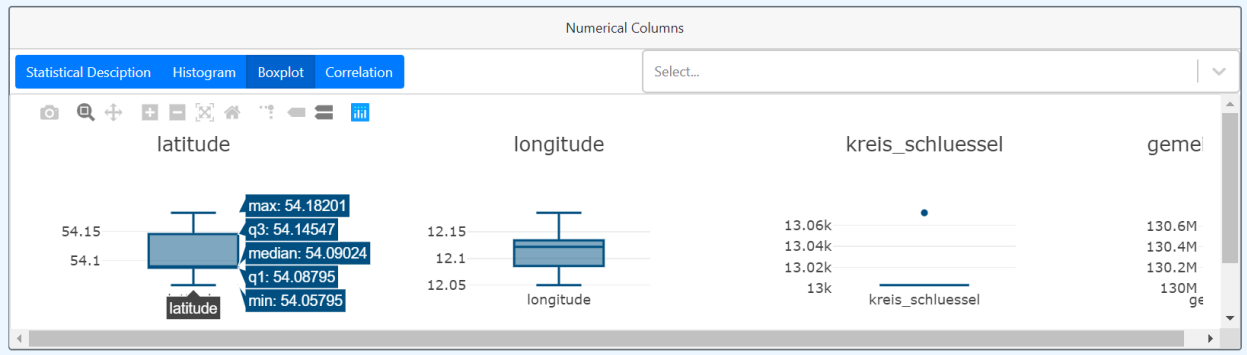
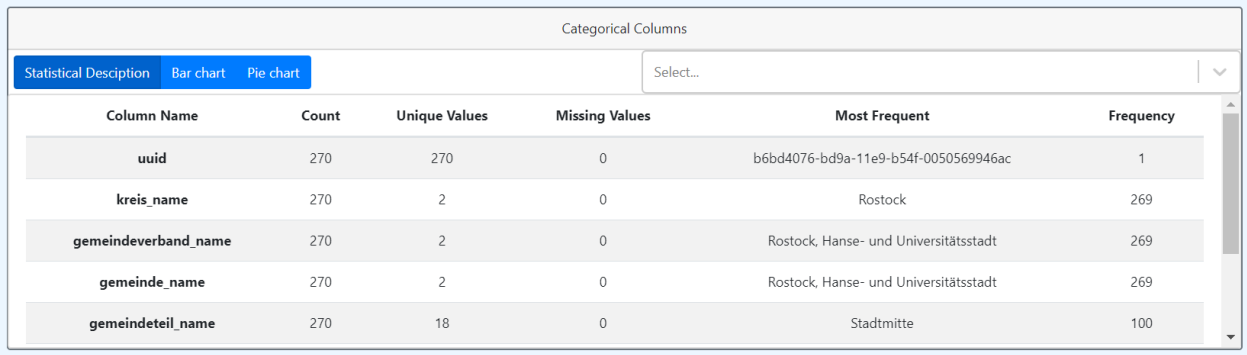


Abb. 20: Visualisierung ausgewählter statistischer Daten der numerischen Spalten des gewählten Datensatzes durch die Nutzerschnittstelle

Abbildung 20 als Ergänzung zu Abbildung 19 stellt zu ausgewählten statistischen Daten die vorgefertigten Visualisierungen in Form Boxplots, Histogrammen oder auch Korrelationsdiagrammen dar.



Column Name	Count	Unique Values	Missing Values	Most Frequent	Frequency
uuid	270	270	0	b6bd4076-bd9a-11e9-b54f-0050569946ac	1
kreis_name	270	2	0	Rostock	269
gemeindeverband_name	270	2	0	Rostock, Hanse- und Universitätsstadt	269
gemeinde_name	270	2	0	Rostock, Hanse- und Universitätsstadt	269
gemeindeteil_name	270	18	0	Stadtmitte	100

Abb. 21: Darstellung der Metadaten zu kategorischen Werten zum gewählten Datensatz durch die Nutzerschnittstelle

Abbildungen 21 und 22 beziehen sich auf die Analyse von Spalten mit kategorischen Werten im vorhandenen Datensatz und zeigen einerseits die vorhandenen statistischen Daten sowie die Visualisierung über verschiedene auswählbare Diagrammtypen.

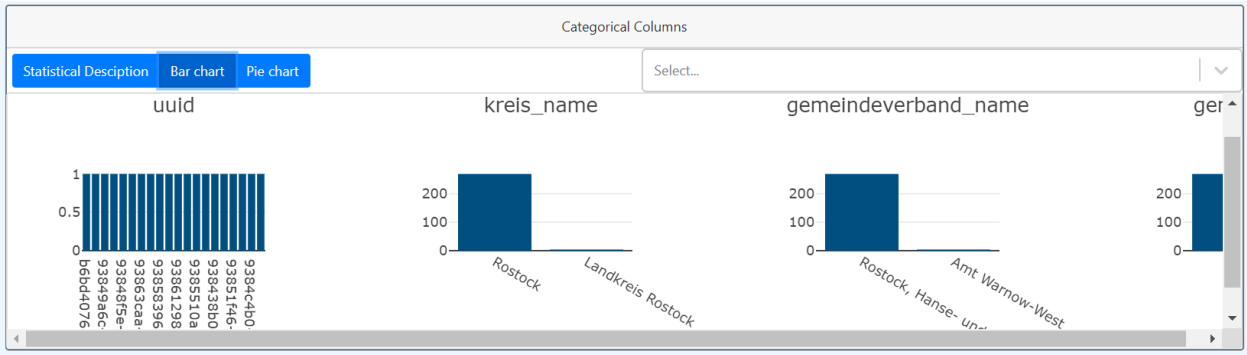


Abb. 22: Visualisierung ausgewählter Metadaten der kategorischen Werte zum gewählten Datensatz durch die Nutzerschnittstelle

Sind im ausgewählten Datensatz räumliche Daten enthalten, so werden sie im Bereich der räumlichen Analyse aufgeführt und über eine einfache Kartendarstellung visualisiert, wie in Abbildung 23 nachfolgend dargestellt.

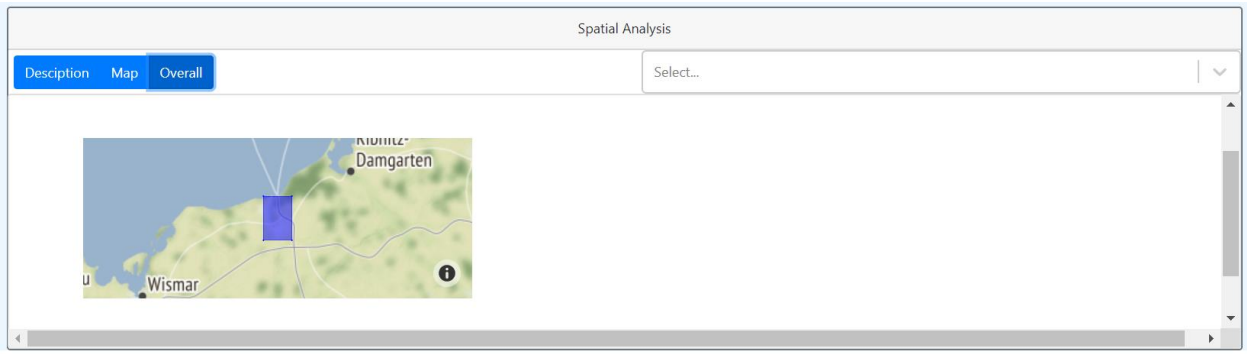


Abb. 23: Visualisierung der Geodaten des gewählten Datensatzes in Form einer Kartendarstellung durch die Nutzerschnittstelle

Die vom Metadatenextraktionstool extrahierten Metadaten werden einerseits für die grafische Nutzerschnittstelle im Datenexplorationstool genutzt, wie oben dargestellt. Der Nutzer kann somit direkt einen Überblick über den Inhalt der jeweiligen Datei gewinnen. Andererseits werden die extrahierten und erzeugten Metadaten von der Entscheidungsmatrix im Demonstrator genutzt, um passende Visualisierungsformen zu empfehlen. Dies wird nachfolgend detailliert vorgestellt.

4.3.2. Demonstrator

Zur Evaluierung der entwickelten Ansätze zur Auswahl von Visualisierungsformen, wurde im Rahmen von mVIZ ein entsprechender Demonstrator konzipiert und implementiert. Dieser Demonstrator verknüpft über eine Benutzeroberfläche die mCLOUD als Katalog mit Werkzeugen zur Metadaten-Extraktion und einer Implementierung der im Projekt entwickelten Entscheidungsmatrix.

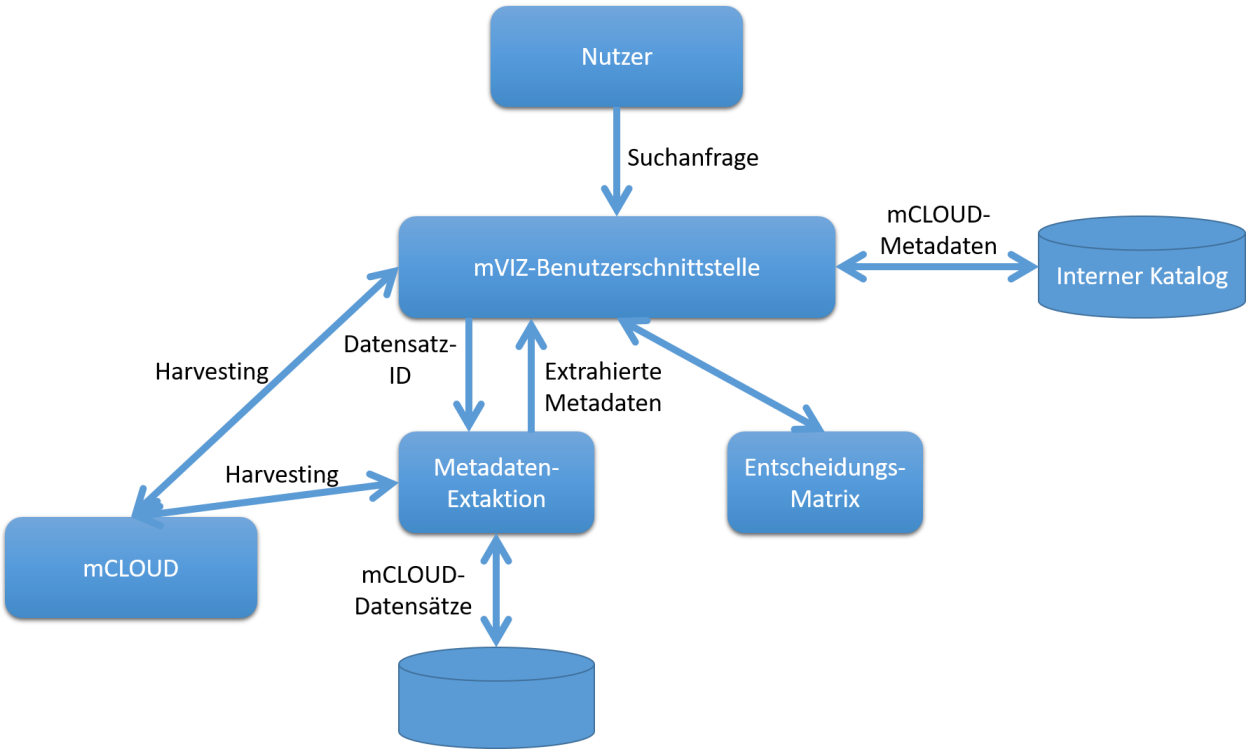


Abb. 24: Workflow des Demonstrators

Abb. 24 stellt den vom Demonstrator implementierten Workflow dar. Zentrales Element ist dabei die mVIZ-Benutzerschnittstelle und die dazugehörende Geschäftslogik. Wird der mVIZ-Demonstrator initialisiert, so erfolgt zunächst ein Sammeln der verfügbaren Metadaten aus der mCLOUD (Harvesting) zur weiteren internen Verwendung. Zu diesem Zweck werden die gesammelten Metadaten in einem (in Memory) Katalog abgelegt. Stellt ein Nutzer eine Anfrage nach bestimmten Daten, so wird diese Suchanfrage zunächst über den internen Katalog beantwortet, so dass der Nutzer ein entsprechendes Suchergebnis erhält (vgl. Abbildung 25). Als Suchfunktionen werden im Rahmen des Demonstrators einerseits eine Textsuche über die Datensatztitel und -beschreibungen und andererseits ein Filter über die Datentypen angeboten.

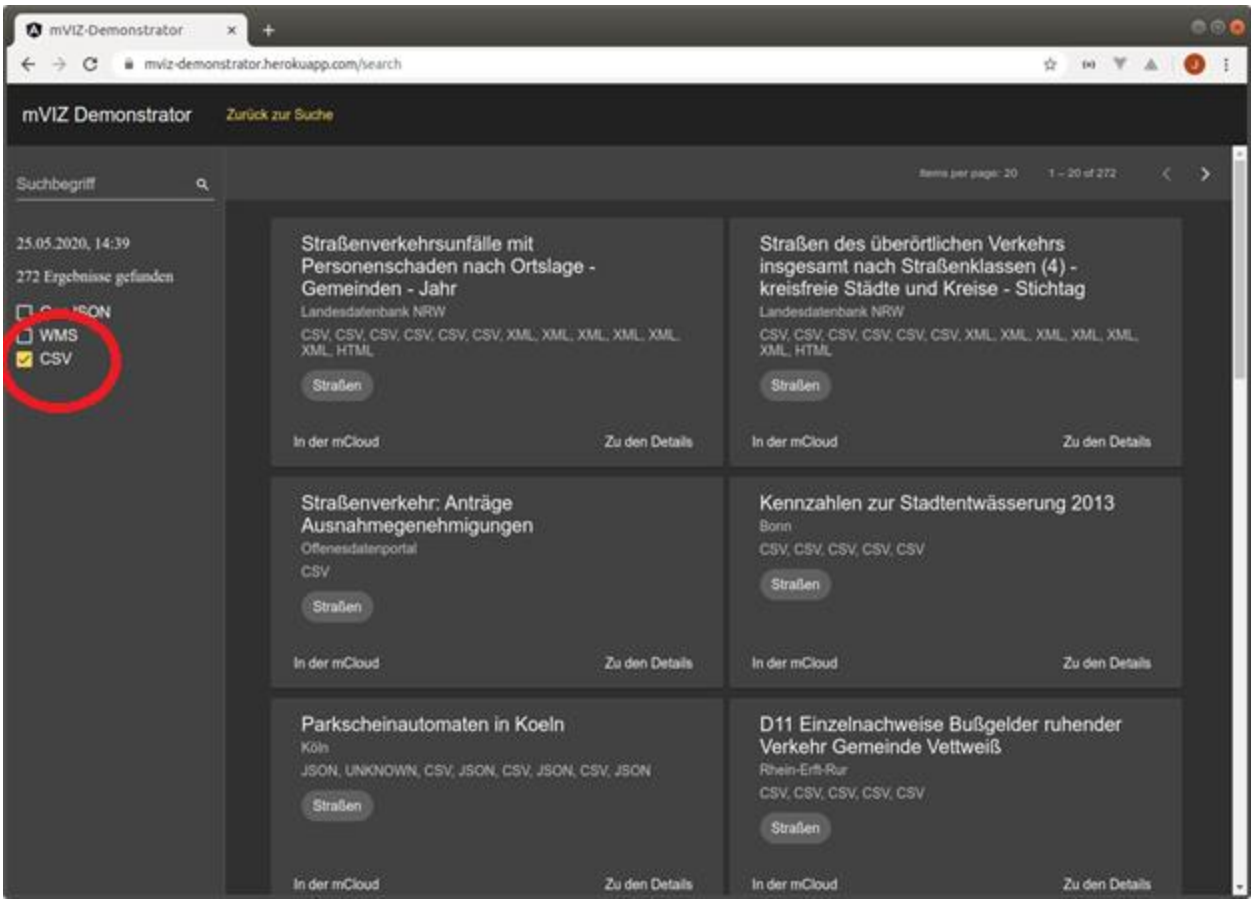


Abb. 25: Suche mit CSV-Parameter als Einschränkung

Von der Übersicht der Suchergebnisse aus stehen weitere Funktionen bereit, welche auf den Forschungsergebnissen des mVIZ-Projekts aufbauen. Einstiegspunkt hierfür ist die Detailansicht der einzelnen Suchtreffer (vgl. Abbildung 25). Von hier aus stehen dem Nutzer verschiedene Optionen zur Verfügung:

- Betrachtung der Metadaten und Absprung in die mCLOUD
- Im Falle von GeoJSON- und WMS-Datenquellen steht eine Preview-Funktion bereit; im Falle von WMS-Datenquellen ist keine weitere Entscheidungsunterstützung zur Visualisierung der gelieferten Daten nötig, da WMS-Server bereits Datenvisualisierungen und nicht die Roh-Daten liefern; im Falle von GeoJSON-Daten beschränkt sich dieser Demonstrator zunächst auf eine Kartendarstellung der enthaltenen Geodaten (Features).
- Liegen CSV-Dateien vor, so wird die mVIZ-Funktionalität zur Auswahl geeigneter Visualisierungen angeboten.

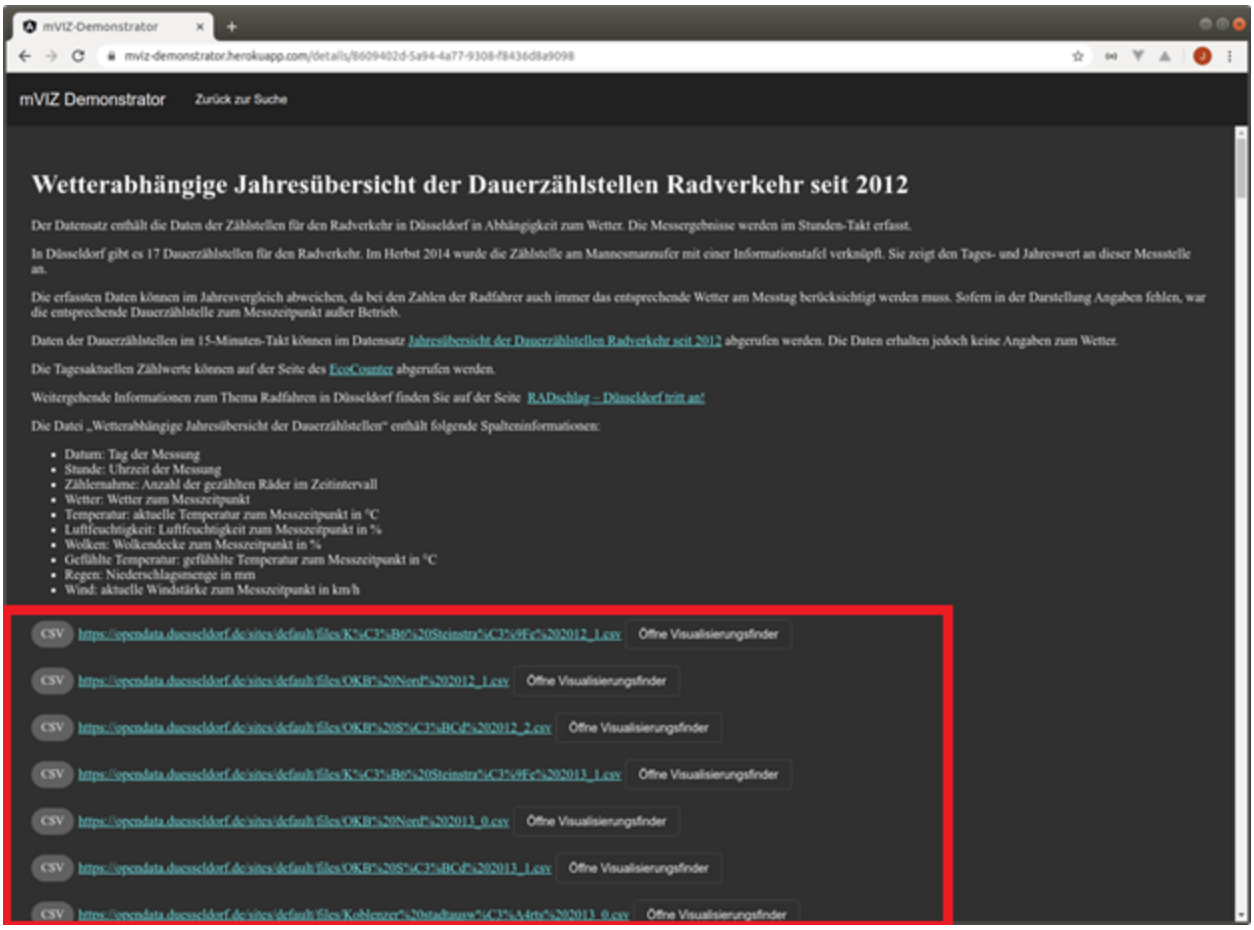


Abb. 26: Auflistung der Dateien zum ausgewählten Datensatz

Handelt es sich um einen CSV-basierten Datensatz, so spricht der mVIZ-Demonstrator das von der Beuth-Hochschule entwickelte Tool zur Metadaten-Extraktion an. Hierbei wird die in der mCLOUD genutzte ID des Datensatzes als Abfrageparameter übergeben. Das Tool zur Metadaten-Extraktion liefert darauf hin alle verfügbaren Metadaten die über den Datensatz gewonnen werden konnten zurück. Diese Metadaten werden nachfolgend verwendet, um die Entscheidungsmatrix zur Visualisierungsauswahl zu parametrisieren (vgl. Abbildung 27).

Anhand der Inputs zu den einzelnen Kriterien der Entscheidungsmatrix wird für jede Visualisierungsmöglichkeit des gewählten Datensatzes ein Scoring berechnet. Anhand dieser Bewertung werden dann dem Nutzer die passendsten Visualisierungsmöglichkeiten vorgeschlagen.

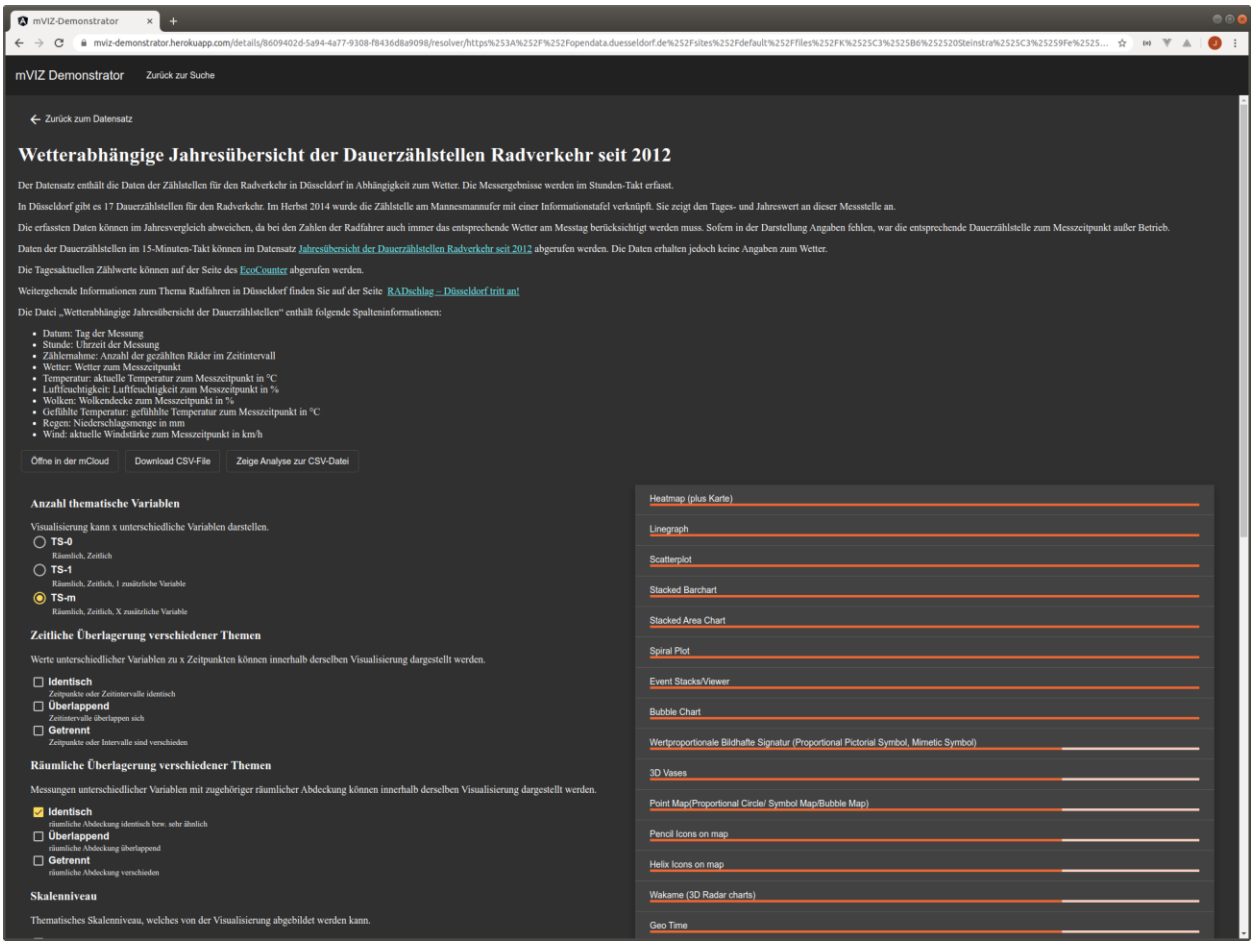


Abb. 27: Visualisierungsauswahl basierend auf der Entscheidungsmatrix

In der Übersicht der potentiell passenden Visualisierungen kann der Nutzer durch einen Klick auf die einzelnen Optionen weitere Informationen erhalten. Dies ist insbesondere eine Illustration, in welcher Form die jeweiligen Visualisierungsansätze dargestellt werden (vergleiche Abbildung 28).

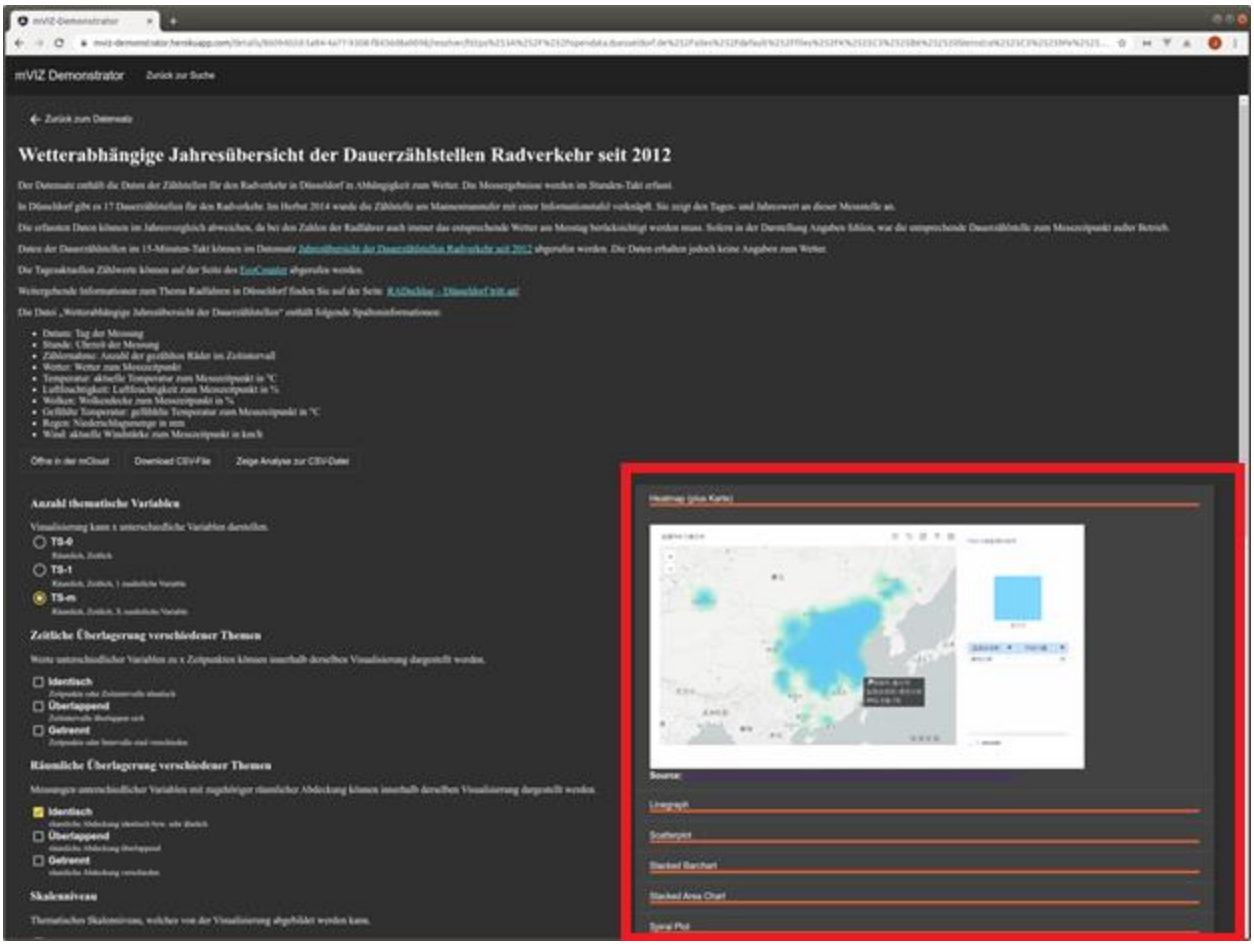


Abb. 28: Visualisierungsauswahl basierend auf der Entscheidungsmatrix mit Zusatzinformation einer Visualisierung

5 Auswertung

In dem vorliegenden Leitfaden wurde ein Konzept vorgestellt, wie basierend auf einer Analyse der Metadaten und Daten eines Datensatzes geeignete Visualisierungen automatisch ausgewählt werden können. Ziel ist es, Nutzern von Open-Data-Portalen, wie der mCLOUD, eine Entscheidungsunterstützung zu bieten und eine möglichst einfache und schnelle Visualisierung der Daten zur effizienten und benutzerfreundlichen Exploration zu ermöglichen.

Fokus des Konzeptes sind zum aktuellen Zeitpunkt raumzeitliche Daten, sowie eine Konzentration auf .csv Dateien.

5.1. Diskussion

Der Leitfaden in seiner aktuellen Version dient dazu, Denkprozesse anzustoßen. Zum gegebenen Zeitpunkt ist die automatische Analyse der Datensätze aus verschiedenen Gründen nur bedingt durchführbar. Beispielsweise erschwert die stark variierende Aufbereitung der Daten durch die Datenproduzenten die Automatisierung. Eine unvollständige Bewertung dazu kann in nachfolgendem Kapitel gefunden werden. Das führt dazu, dass im vorliegenden Leitfaden vorrangig Empfehlungen ausgesprochen werden, wie Datensätze in Bezug auf Metadaten und Beschreibungen aussehen könnten, um eine Analyse und automatische Auswahl von Visualisierungen zu ermöglichen.

Mit dem Kern des Konzeptes wurde eine Liste von Kriterien erarbeitet, welche zur Auswahl geeigneter Visualisierungen notwendig sind. Diese dienen dazu, den Datenproduzenten und Entwicklern ein Gespür dafür zu vermitteln, welche Angaben für ein solches System essentiell wären und wie Open-Data-Angebote, wie die mCLOUD, zugänglicher gemacht werden können.

Durch die Fokussierung auf raumzeitliche Daten beschreibt das Konzept nur einen kleinen Bereich der großen Menge an verfügbaren Daten. Eine ganzheitliche Analyse sämtlicher Bereiche ist zum jetzigen Zeitpunkt nicht denkbar, dennoch liefert die Visualisierungsmatrix erste Ideen, für eine umfassende Erweiterung auf andere Bereiche.

Eine weitere Herausforderung ist es, einen geeigneten Feedback-Mechanismus zu finden, der es ermöglicht, das Konzept zu evaluieren und valide Aussagen darüber zu treffen, inwiefern eine vorgeschlagene Visualisierung auf einen Datensatz passt. Zum aktuellen Projektstand konnte diese Evaluierung auf Grund der globalen Situation durch die Corona-Pandemie nur durch vereinzelte Experten überprüft werden.

Mit mVIZ wurde versucht eine Gratwanderung zwischen Nutzeranforderungen (wie z.B.: einfachen Visualisierungen) und Komplexität durch den Charakter der Daten zu schaffen. Dabei wurden einige Visualisierungen als ungeeignet bewertet, die von einem professionellen Visualization Designer durchaus genutzt werden könnten. Da unsere Zielgruppe sich jedoch aus

einer Vielzahl von Fachbereichen zusammensetzt, wurde auf eine komplexere Auswertung verzichtet.
Für manche Daten können sich Visualisierungen gegebenenfalls auch direkt aus den verfügbaren Datenformaten ergeben, so dass eine umfangreiche Analyse gar nicht erst nötig wäre und die Funktion einer Visualisierung für die Vorschau genutzt werden kann. Im Demonstrator von mVIZ ist dies für GeoJSON exemplarisch umgesetzt, es wäre jedoch auch denkbar, dies für viele fachspezifische Formate in ähnlicher Weise anzubieten (z.B. NetCDF-Dateien, Datex II-Daten, GRIB-Dateien).

Nichtsdestotrotz bleibt eine vollautomatische Auswahl ohne Einbeziehung der Nutzer schwierig. Viele Entscheidungen bezüglich der Visualisierung hängen von individuellen Präferenzen, sowie individuellen Zielen und Nutzungskontexten ab. Dennoch sollte es durch mVIZ möglich sein, Entscheidungen zu unterstützen und Nutzern dabei zu helfen, eine erste Analyse verfügbarer Datensätze durchzuführen.

5.2. Metadaten Schema

Bei der Untersuchung der Datenpublikationen der mCLOUD, die sich auf CSV-Dateien als dem am häufigsten vorkommenden Dateityp innerhalb der Datenpublikationen fokussierte, wurden immer wiederkehrende Probleme identifiziert. Diese stehen durchaus für symptomatische Fehlerarten in Open-Data-Portalen am Beispiel der mCLOUD. Nachfolgend werden die verschiedenen Problemklassen beschrieben und mit Beispielen illustriert und, wenn möglich, Empfehlungen zur Verbesserung aufgezeigt.

Problem	Beschreibung	Erläuterung / Beispiel
	Beschreibung / textuelles Beispiel	Besonders bei größeren Dateien kommt es vor, dass die Datei in einem Zip-Ordner eingepackt ist, ohne dies in der RDF Metadatenbeschreibung als zip gekennzeichnet zu haben
	Verbesserungs-empfehlung	Die Komprimierung sollte in den Metadaten erwähnt werden.
	Beispiel aus der mCLOUD	https://mcloud.de/web/guest/suche/-/results/detail/5000642e-f686-4211-b7c7-97a9c2c1e23a
Zip Ordner beinhaltet	Beschreibung / textuelles Beispiel.	Jeder Datensatz der mCLOUD darf mehrere Distributionen haben, wenn jede dieser Distributionen auch mehrere Dateien beinhalten

mehr als eine CSV-Datei		würde, steigt die Komplexität des Systems unnötig und kann zur Verwirrung führen.
	Verbesserungs-empfehlung	Es sollte eine Kollektion der Dateien von einzelnen Dateien unterschieden werden.
	Beispiel aus der mCLOUD	https://mcloud.de/web/guest/suche/-/results/detail/d1e9afdf-fbf5-482e-aecd-9777225a03f7
Falsch eingeordnet	Beschreibung / textuelles Beispiel	Nicht CSV-Distributionen werden fälschlich als CSV in der mCLOUD verzeichnet.
	Verbesserungs-empfehlung	Fehlerkorrektur
	Beispiel aus der mCLOUD	https://mcloud.de/web/guest/suche/-/results/detail/0d8b65c3-05c1-4d6f-bc42-635e75f3e88b
Datei beinhaltet fehlerhafte Zeichen	Beschreibung / textuelles Beispiel	Zum Teil treten unzulässige Zeichen in Dateien auf, wie z.B. Stra§e, BahnŸberfŸhrung etc.
	Verbesserungs-empfehlung	Für besondere deutsche Buchstaben wie Umlaute und “ß” sollte entweder die Schreibung über AE, UE usw. genutzt werden oder es sollte beim Export darauf geachtet werden, dass die Codierung korrekt erfolgt.
	Beispiel aus der mCLOUD	https://mcloud.de/web/guest/suche/-/results/detail/acde1b43-d6ce-434c-bf2b-aec43bf3d815
Unnötige Verwendung von “ ”	Beschreibung / textuelles Beispiel	Teilweise werden unnötige Aneinanderreihungen von “ “ vorgenommen, wie """2011""".
	Verbesserungs-empfehlung	Anführungszeichen sollten wohldosiert eingesetzt werden.

	Beispiel aus der mCLOUD	https://mcloud.de/web/guest/suche/-/results/detail/d02e9706-af6b-4f53-8681-e11b3e74c562
Verwendung von Einheiten in Tabellen-zellen	Beschreibung / textuelles Beispiel	Teilweise werden Mengen- und Maßeinheiten in Zusammenhang mit numerischen Werten angegeben, wie z.B. 20 € anstelle einer Reduzierung auf den numerischen Wert 20. Damit werden alle andere Datentypen (Zahlen, usw.) als String interpretiert.
	Verbesserungs-empfehlung	Einheiten sollten in den Überschriften oder in begleitenden Metadaten erwähnt werden.
	Beispiel aus der mCLOUD	https://mcloud.de/web/guest/suche/-/results/detail/a5413176-635e-4b18-9e2c-abf3a3237de2
Unangemessener Einsatz von Trenn-zeichen	Beschreibung / textuelles Beispiel	Leerzeichen sind für CSV Dateien mit textuellen Inhalten unangemessen. Jede Instanz eines Trennzeichens wird als eine neue Spalte interpretiert.
	Verbesserungs-empfehlung	Je nach Inhalt sollte eines der Standard Trennzeichen verwendet werden (, ; \t).
	Beispiel aus der mCLOUD	Unangemessene Trennzeichen
Unerwünschte Texte am Anfang oder / und am Ende	Beschreibung / textuelles Beispiel	Die Datei Beschreibung ist mitunter zu Beginn oder am Ende der Tabelle vermerkt und sollte nicht in die Datei geschrieben werden.
	Verbesserungs-empfehlung	CSV Dateien sollten nur mit tabellarischen Werten befüllt werden.
	Beispiel aus der mCLOUD	https://mcloud.de/web/guest/suche/-/results/detail/ce5916e8-f332-57f7-9d33-48d47cf9e39c

Mehrere Zeilen von Überschriften	Beschreibung / textuelles Beispiel	In vielen CSV Verarbeitungsbibliotheken werden die Zeilen nach der ersten Zeile benutzt, um die Datentypen zu bestimmen.
	Verbesserungs-empfehlung	Nur eine Zeile sollte den Überschriften gewidmet werden.
	Beispiel aus der mCLOUD	https://mcloud.de/web/guest/suche/-/results/detail/5741c093-2747-5412-96be-22c0a0913934
Nicht eingehaltene tabellarische Struktur	Beschreibung / textuelles Beispiel	Das CSV-Schema ist für die Speicherung von tabellarischen Daten gedacht. Alle anderen dort vorkommenden Inhalte widersprechen diesem Grundkonzept.
	Verbesserungs-empfehlung	Die tabellarische Struktur sollte eingehalten werden.
	Beispiel aus der mCLOUD	https://mcloud.de/web/guest/suche/-/results/detail/364fda21-fb6a-4415-a12c-9b876114bb85
Leere Zeilen / Spalten	Beschreibung / textuelles Beispiel	In CSV Dateien werden leere Zeilen oder Spalten nicht ignoriert, sondern derartige Einträge als „fehlende Werte“ interpretiert. Deswegen dürfen Lücken nicht für die Verbesserung der Lesbarkeit verwendet werden.
	Verbesserungs-empfehlung	Leere Zeilen bzw. Spalten sollten somit vermieden werden.
	Beispiel aus der mCLOUD	https://mcloud.de/web/guest/suche/-/results/detail/CE5E73DD-E0A1-4EE6-8C31-A52607B6502F

Mangelnde oder nicht beschreibende Überschriften	Beschreibung / textuelles Beispiel	Obwohl das Vorhandensein von Überschriften nicht nötig ist, wäre deren obligatorische Präsenz sehr hilfreich. In bestimmten Fällen können die Überschriften in die Analyse miteinbezogen werden, um bessere Ergebnisse zu erzielen. Ein Beispiel dafür sind die Angaben der geografischen Koordinaten. Wenn die Daten sich auf Regionen beziehen, deren Koordinaten zwischen 0 und 90 liegen (wie für Deutschland korrekt), ist es unmöglich, Latitude und Longitude ohne passende Überschriften zu unterscheiden.
	Verbesserungs-empfehlung	Es sollten immer gut beschreibende Überschriften verwendet werden.
Uneinheitliche Schreibweise	Beschreibung / textuelles Beispiel	Mitunter werden deutsche oder englische (usw.) Schreibweisen gemischt verwendet. Beispiele hierfür wären: in einer Spalte werden die Dezimalzahlen mit Punkten gekennzeichnet, in anderen Spalten mit Kommas.
	Verbesserungs-empfehlung	Generell, soweit es möglich ist, sollte eine einheitliche Schreibweise im Dokument verwendet werden.

Tabelle 17: Metadatenprobleme und Lösungsvorschläge

Neben den vorgefundenen Fehlerarten in den CSV-Dateien der mCLOUD kann insgesamt als problematisch für dieses Projekt zusammengefasst werden:

- Die üblichen Metadaten-Standards bieten nicht genug Felder für die Umsetzung des Konzepts im Projekt mVIZ. Insofern sollten die Metadaten-Schemata weiter dokumentiert und mit offiziellen Metadatenstandards abgeglichen werden (ISO, DCAT-AP). Möglicherweise müssen ggf. entsprechende Erweiterungen formal definiert werden.
- Eine breitere Analyse müsste über verschiedene Anwendungsdomänen hinweg durchgeführt werden, die die folgenden Fragestellungen adressiert: Welche Informationen bieten die dort verfügbaren Metadatenstandards/Datenformate? Können hier ggf. einzelne Ansätze für einen gemeingültigen Metadaten-Ansatz übernommen werden?
- Eine Weiterentwicklung eines gemeinsamen, internen Datenformats als Input für Vorschauvisualisierungen ist ggf. erforderlich.

6 Ausblick

Aus den Erkenntnissen, welche im Rahmen des als Vorstudie konzipierten mVIZ-Projekts gewonnen wurden, konnten zahlreiche Handlungsempfehlungen für eine mögliche Weiterführung abgeleitet werden. Diese sollten Bestandteil einer anschließenden Projektphase (mVIZ 2.0) sein, welche jedoch aufgrund der aktuellen Bedingungen voraussichtlich nicht im Rahmen des mFUND durchgeführt werden kann. Dies betrifft eine Reihe verschiedener Handlungsfelder, welche nachfolgend vorgestellt werden:

6.1. Erweiterung der Empfehlungskriterien

Im Rahmen von mVIZ wurde ein umfassender Katalog von Kriterien entwickelt, anhand derer aus den Eigenschaften eines Datensatzes Empfehlungen für geeignete Visualisierungen abgeleitet werden können. Bei der Evaluierung und im nachfolgenden fachlichen Diskurs sind jedoch noch weitere mögliche Kriterien identifiziert worden, die für eine Weiterentwicklung des Kriterienkatalogs zu prüfen sind. Dies sind beispielsweise:

- Zur Verfügung stehende Anzeigefläche (werden kompakte oder umfangreiche Visualisierungen benötigt),
- Verständlichkeit für den Nutzer (z.B. spezifische Darstellung für Fachanwender; einfache Darstellung für die breite Öffentlichkeit)
- Komplexität der Umsetzung einer Visualisierung
- Themenspezifische Datensatzeignung
- Expressivität
- Effektivität

Darüber hinaus ist zu prüfen, wie eine Gewichtung der Auswahlkriterien und die Definition von Ausschlusskriterien zu einer weiteren Verbesserung der Visualisierungsvorschläge führen kann.

6.2. Metadaten-Extraktion

Die Möglichkeit der Extraktion von Metadaten aus mCLOUD-Datensätzen zur Beantwortung der Entscheidungskriterien für eine Visualisierungsauswahl wurde in mVIZ als technisch möglich demonstriert. Allerdings haben diese Arbeiten den Charakter eines ersten Machbarkeitsnachweises. Daher sind auf diesem Gebiet noch tiefergehende Arbeiten erforderlich. Dies sind insbesondere:

- Umfassende Analyse, zu welchen Entscheidungskriterien eine Metadatenextraktion möglich ist
- Systematische Umsetzung der Metadaten-Extraktion für alle Felder, bei denen eine Extraktion sinnvoll machbar ist
- Evaluierung der Zuverlässigkeit der Metadaten-Extraktion

- Analyse halbautomatischer Extraktions-Workflows für Fälle in denen eine vollautomatische Metadatenextraktion nicht möglich ist (z.B. Wizards zur Gewinnung zusätzlicher Informationen vom Anwender)
- Ausweitung auf weitere Datenformate über CSV hinaus
- Einbeziehung impliziter Metadaten, welche sich aus dem Vorhandensein bestimmter Dateiformate ergeben

6.3. Feedback-Mechanismus und Einbeziehung von Machine Learning-Methoden

Im aktuellen Demonstrator wurde noch kein Feedback-Mechanismus vorgesehen, mit dem Nutzer eine Rückmeldung geben können, inwiefern die vorgeschlagenen Visualisierungsformen tatsächlich für die jeweiligen Datensätze geeignet sind. Ein solcher Mechanismus bietet ein großes Potential in zweierlei Hinsicht:

Einerseits können über die Rückmeldung von Nutzern ggf. die Entscheidungskriterien erweitert, ihre Gewichtung modifiziert und Entscheidungswege verbessert werden. Dies käme der Weiterentwicklung der Entscheidungsmatrix und ihrer Auswertung zu Gute.

Andererseits bieten die aktuell immer weiter verbreiteten Lösungen aus dem Bereich des maschinellen Lernens weitere interessante Perspektiven. Hier wäre zu untersuchen, inwieweit die Entscheidungsmatrix hierdurch ergänzt werden könnte. Beispielsweise könnte das Feedback der Nutzer verwendet werden, um einen entsprechenden Vorschlagsalgorithmus zu trainieren.

6.4. Weiterentwicklung des Demonstrators

Der in mVIZ umgesetzte Demonstrator wurde mit dem Ziel entwickelt, die technische Machbarkeit der Entscheidungsunterstützung bei der Auswahl geeigneter Visualisierungen für mCLOUD-Datensätze nachzuweisen. Auf dieser Basis bietet sich die Weiterentwicklung zu einem Prototyp an. Dies würde verschiedene Arbeiten im Hinblick auf die Usability und den Funktionsumfang umfassen.

Ein wichtiger Baustein wäre die Bereitstellung von Visualisierungsmodulen für alle vom System unterstützen Visualisierungsmethoden. Dies würde die Option eröffnen, dass Nutzer bei der Auswahl von Visualisierungsmethoden eine Vorschau erhalten könnten, wie eine Visualisierung eines Datensatzes bei Nutzung der einzelnen Verfahren aussehen würde. Gleichzeitig könnte diese Funktionalität über eine Plug-In-Architektur realisiert werden, so dass zusätzliche Visualisierungsformen dynamisch im System ergänzt werden könnten. Dies könnte bis hin zu einer Open-Community im mFUND-Umfeld reichen, über die Forschungsprojekte zur Erweiterung der Visualisierungsfähigkeiten beitragen könnten.

Weiterhin ist die Durchführung einer umfassenden Usability-Studie sinnvoll. Im Rahmen von mVIZ konnte dies für den Demonstrator nicht durchgeführt werden. Unter Berücksichtigung der weiter oben vorgeschlagenen Erweiterungen und Arbeiten ergeben sich verschiedene Optionen,

wie ein Nutzer-Workflow gestaltet werden kann. Um hier eine möglichst nachhaltige Entwicklung mit einer Option zur zukünftigen Einbindung in die mCLOUD durchzuführen, empfehlen wir die parallele Begleitung durch entsprechende Nutzerstudien.

Anhang A Liste raumzeitlicher Visualisierungen

3D ICONS AUF KARTEN

DATA VASES

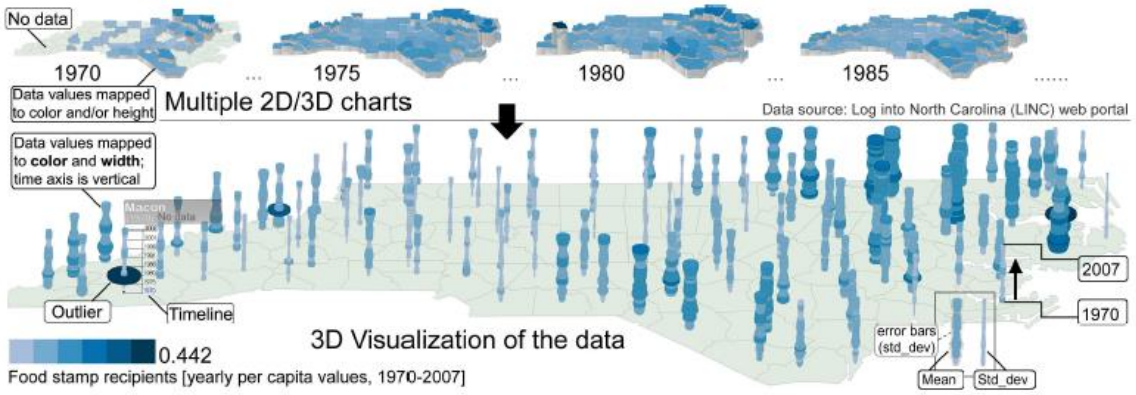


Figure 3. Visualizations of the spatial and temporal distributions of food stamps issued in the 100 counties in North Carolina through a nation-wide program of the federal government of USA. In the standard small multiples [26] display of the data (top) some interesting patterns are visible; for example, higher food stamp claims occur in the north-eastern part of the state. Our 3D visualization on the bottom provides enhanced access to the spatio-temporal distributions of the food stamp claims and selected details of changes in the claims in individual counties, all in a single view.

S. Thakur and A. J. Hanson, "A 3D Visualization of Multiple Time Series on Maps," *2010 14th International Conference Information Visualisation*, London, 2010, pp. 336-343, doi: 10.1109/IV.2010.54.

PENCIL & HELIX ICONS ON MAP

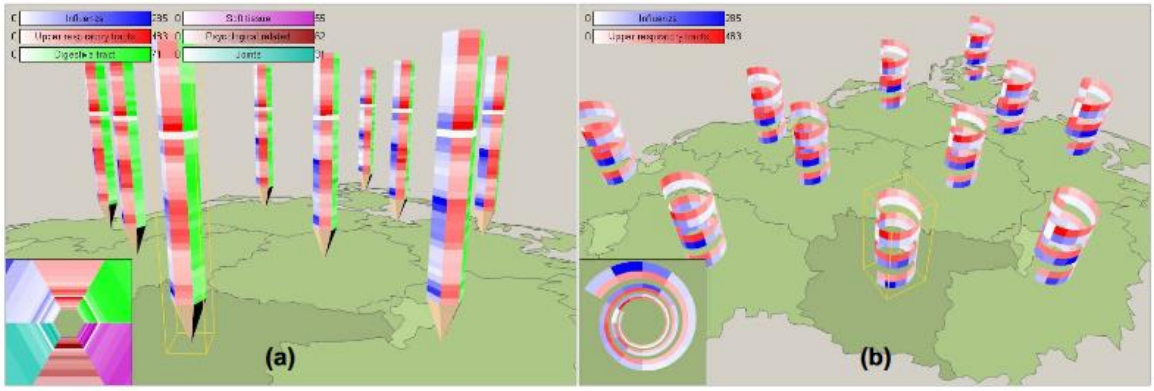
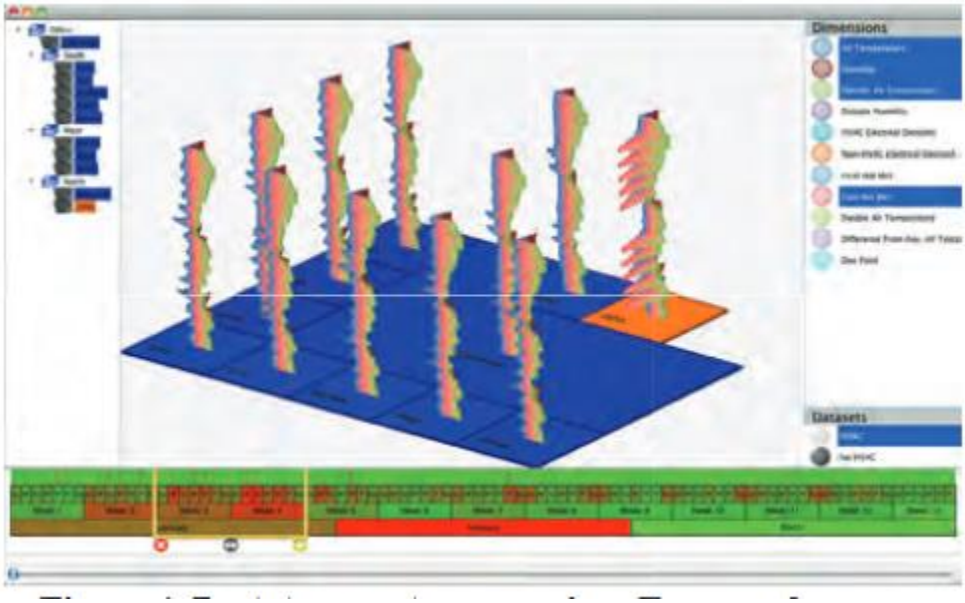


Figure 1. Visualizing monthly health data by means of 3D icons on a map: (a) Pencil icons representing cases of 6 diseases, some diseases show a certain pattern over time; (b) Helix icons clearly reveal the cyclic characteristic of 2 selected diseases. Additional "tunnel views" mitigate the problem of hidden information for a selected icon.

C. Tominski, P. Schulze-Wollgast and H. Schumann, "3D information visualization for time dependent data on maps," *Ninth International Conference on Information Visualisation (IV'05)*, London, UK, 2005, pp. 175-181, doi: 10.1109/IV.2005.3.

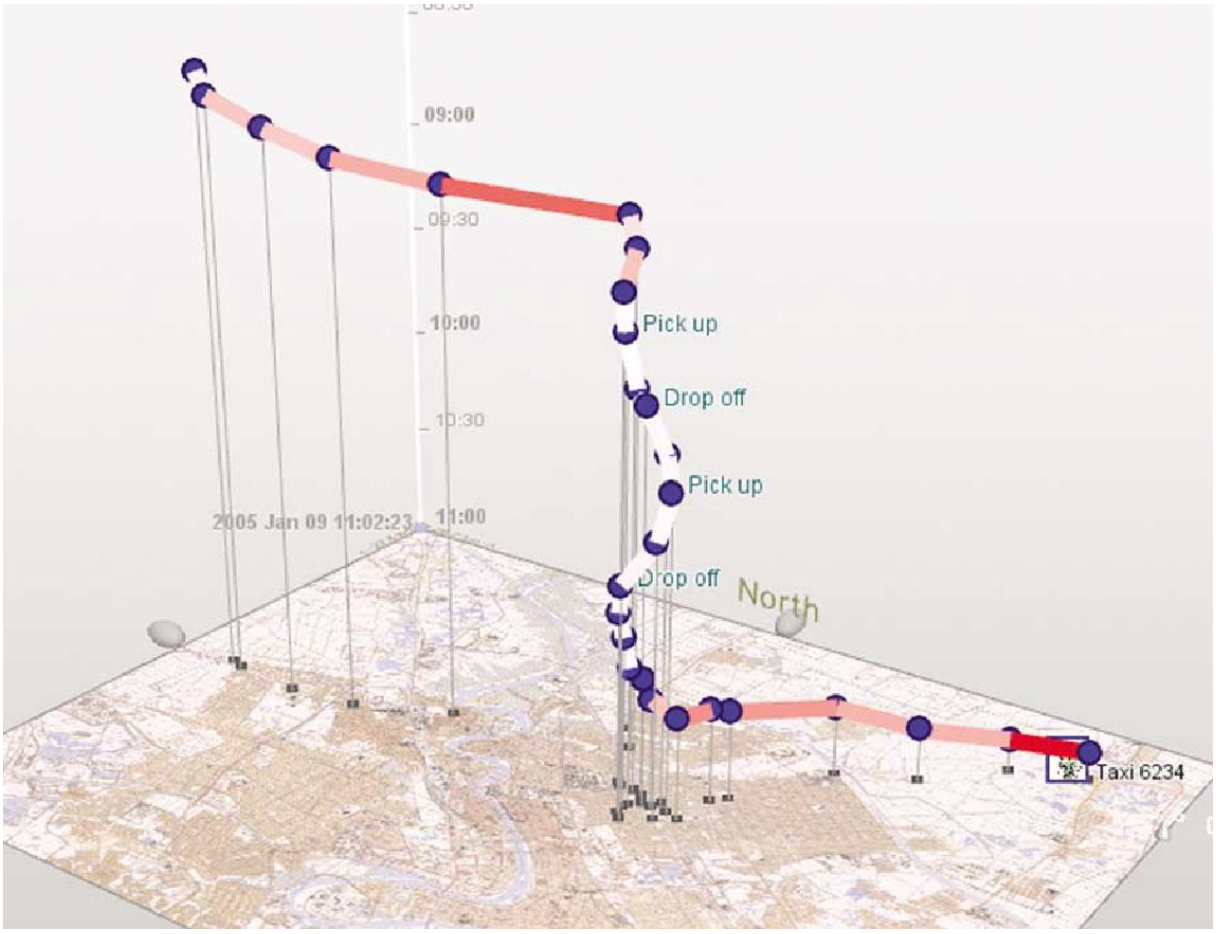
WAKAME



Clifton Forlines and Kent Wittenburg. 2010. Wakame: sense making of multi-dimensional spatial-temporal data. In Proceedings of the International Conference on Advanced Visual Interfaces (AVI '10). Association for Computing Machinery, New York, NY, USA, 33–40. DOI:https://doi.org/10.1145/1842993.1843000

SPACE TIME CUBE & TRAJECTORIES

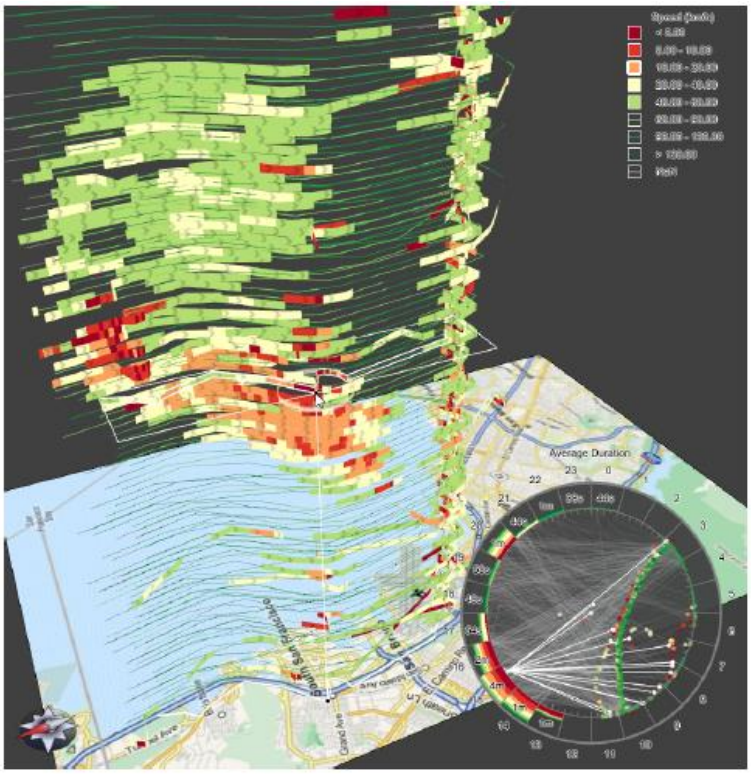
GEOTIME



KAPLER, Thomas; WRIGHT, William. Geotime information visualization. Information visualization, 2005, 4. Jg., Nr. 2, S. 136-146.

ECCLES, Ryan, et al. Stories in GeoTime. In: 2007 IEEE Symposium on Visual Analytics Science and Technology. 2007. S. 19-26.

TRAJECTORY WALL



TOMINSKI, Christian, et al. Stacking-based visualization of trajectory attribute data. IEEE Transactions on visualization and Computer Graphics, 2012, 18. Jg., Nr. 12, S. 2565-2574.

DENSITY TRAJECTORY WALL

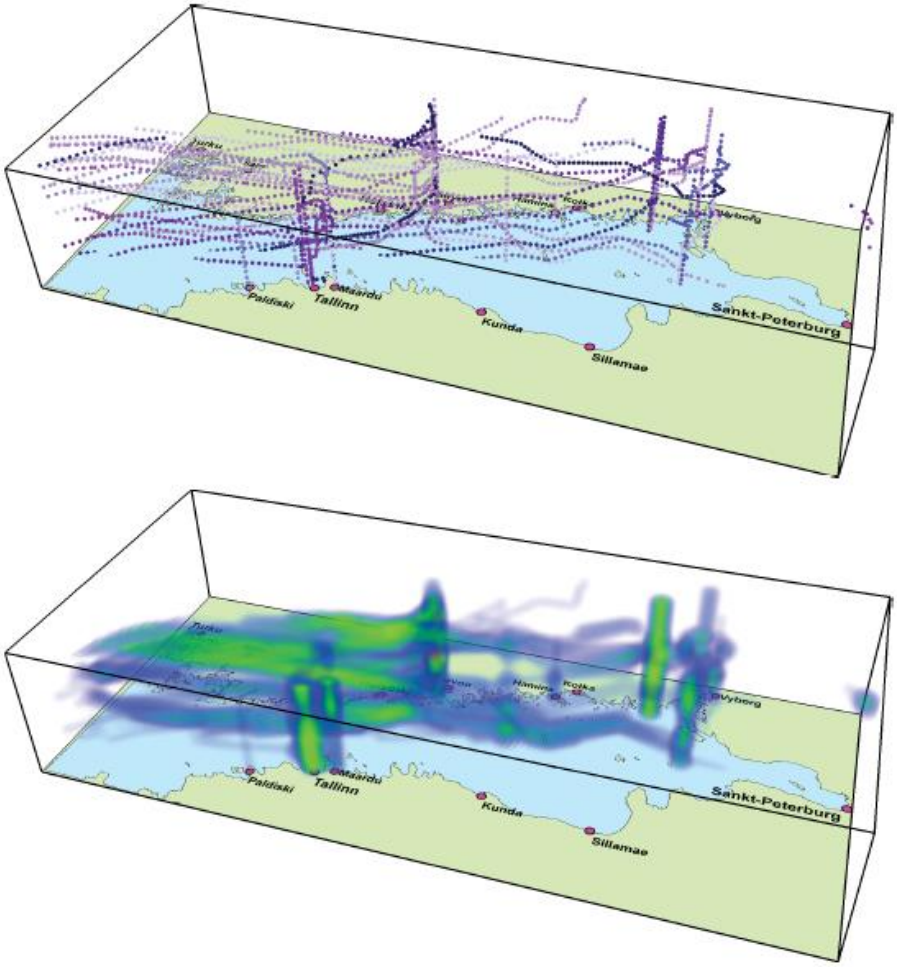
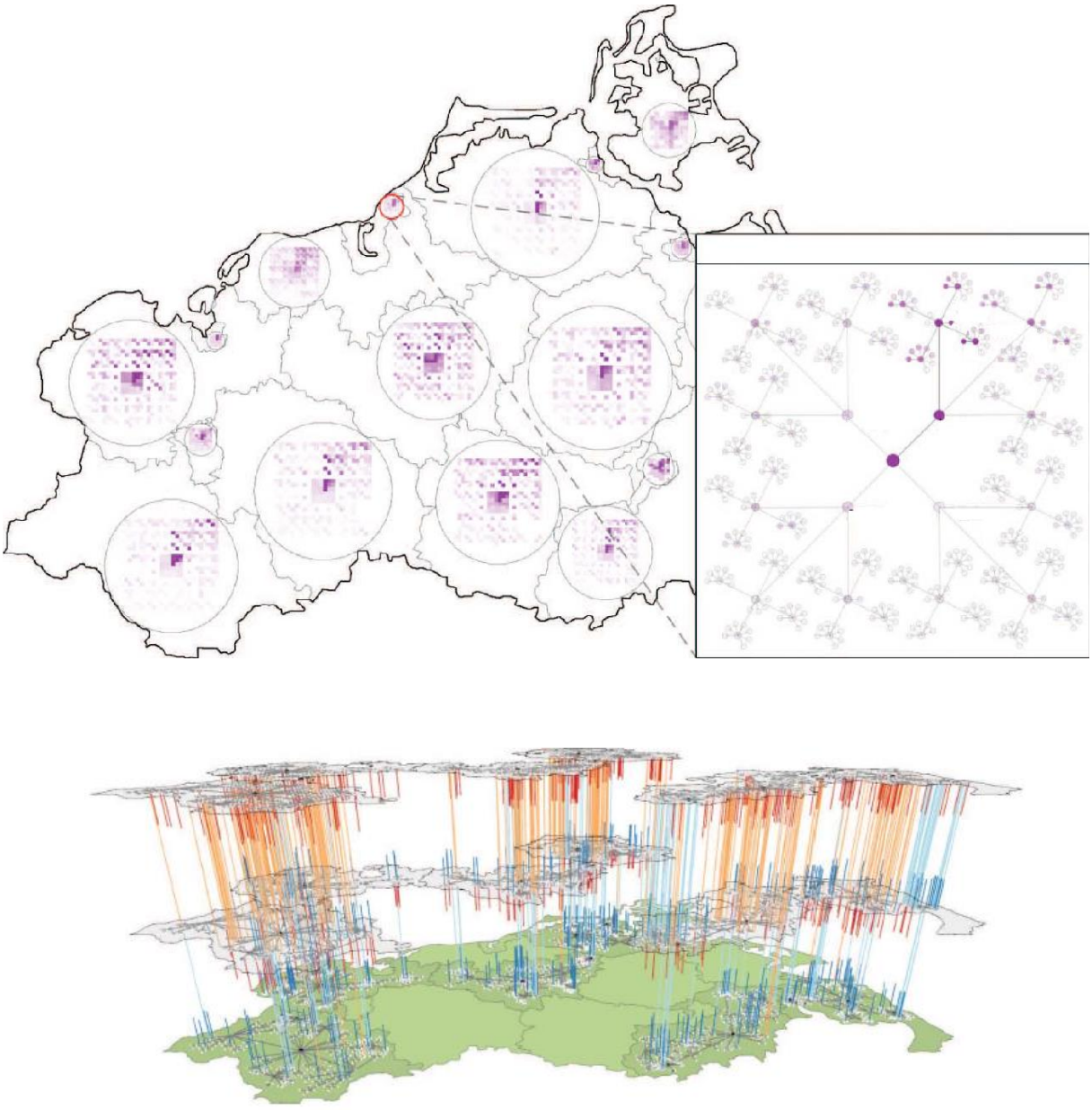


Figure 3. Aggregation of trajectories by computing space–time density of the movement. Top: trajectories of tankers during 1 day are depicted as traces in the space–time cube. Bottom: the space–time density of the tanker movement is shown in the space–time cube using volume-rendering technique.

ANDRIENKO, Gennady, et al. Space, time and visual analytics. International journal of geographical information science, 2010, 24. Jg., Nr. 10, S. 1577-1600.

Urška Demšar & Kirsi Virrantaus (2010) Space–time density of trajectories: exploring spatio-temporal patterns in movement data, International Journal of Geographical Information Science, 24:10, 1527-1542, DOI: 10.1080/13658816.2010.511223



S. Hadlak, C. Tominski, H.-J. Schulz & H. Schumann (2010) Visualization of attributed hierarchical structures in a spatiotemporal context, International Journal of Geographical Information Science, 24:10, 1497-1513, DOI: 10.1080/13658816.2010.510840

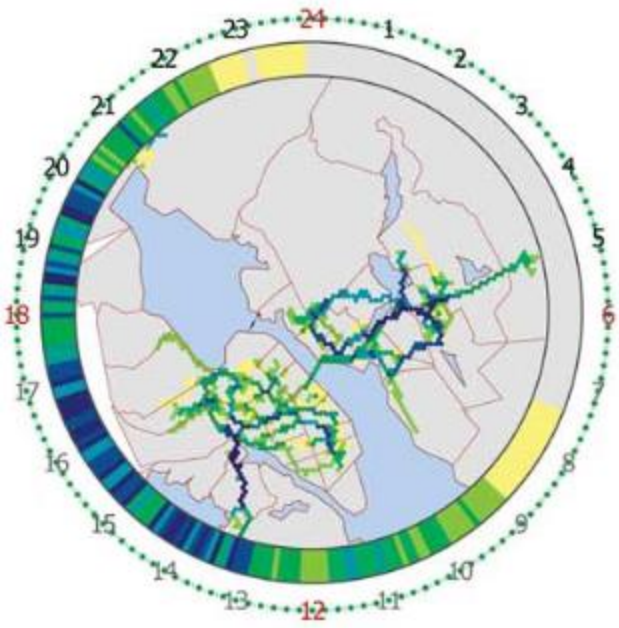
2D KARTENVISUALISIERUNGEN

WERTPROPORTIONALE BILDHAFTE SIGNATUR

SLOCUM, T. A., et al. Thematic cartography and geographic visualization. 2005.

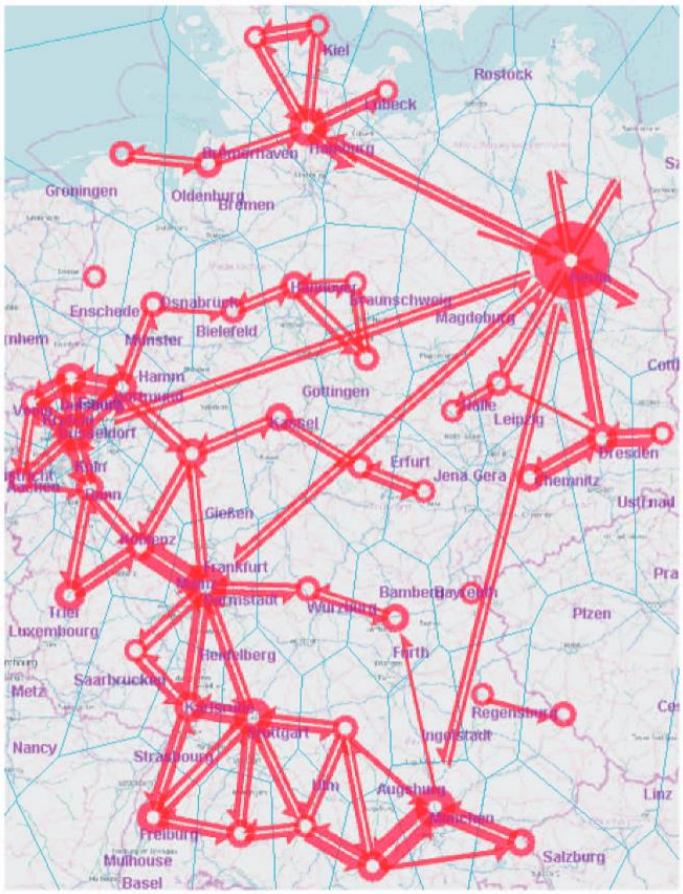
SCHNABEL, O. Benutzerdefinierte diagrammsignaturen in karten. Konzepte, Formalisierung und Implementation. Diss. ETH, 2007, Nr. 16977.

RING MAPS



ZHAO, Jinfeng; FORER, Pip; HARVEY, Andrew S. Activities, ringmaps and geovisualization of large human movement fields. Information visualization, 2008, 7. Jg., Nr. 3-4, S. 198-209.

FLOW MAP



N. Adrienko and G. Adrienko, "Spatial Generalization and Aggregation of Massive Movement Data," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 2, pp. 205-219, Feb. 2011, doi: 10.1109/TVCG.2010.44.

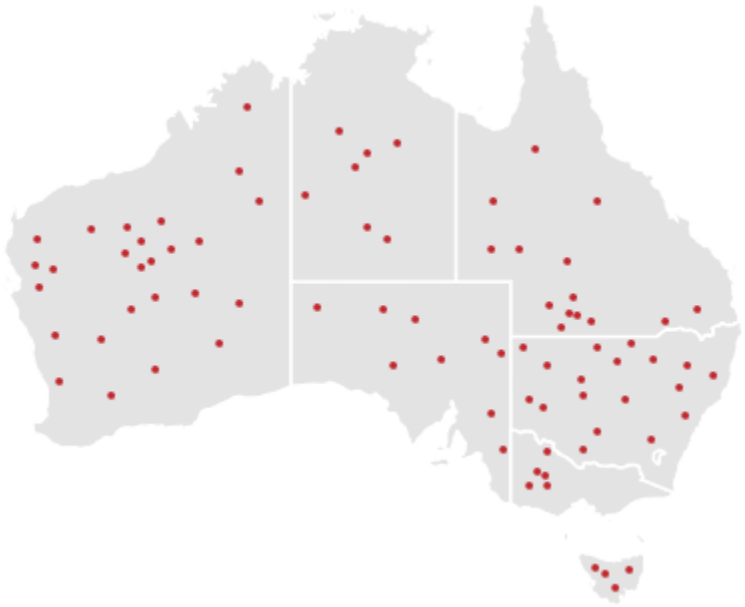
Line Map

Lewis Chou, „Top 10 Map Types in Data Visualization“, Medium, 26. Mai 2020, https://towardsdatascience.com/top-10-map-types-in-data-visualization-b3a80898ea70

Point Map

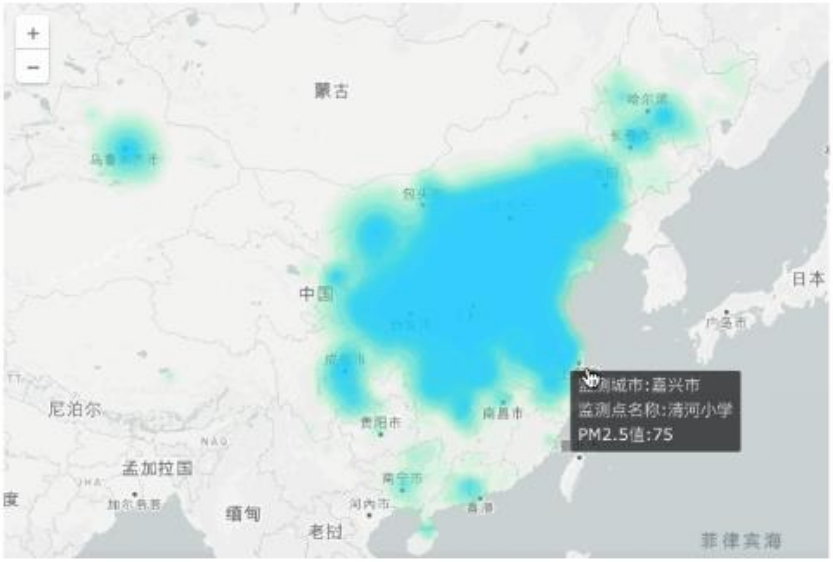
Lewis Chou, „Top 10 Map Types in Data Visualization“, Medium, 26. Mai 2020, https://towardsdatascience.com/top-10-map-types-in-data-visualization-b3a80898ea70

Dot Densitiy Map



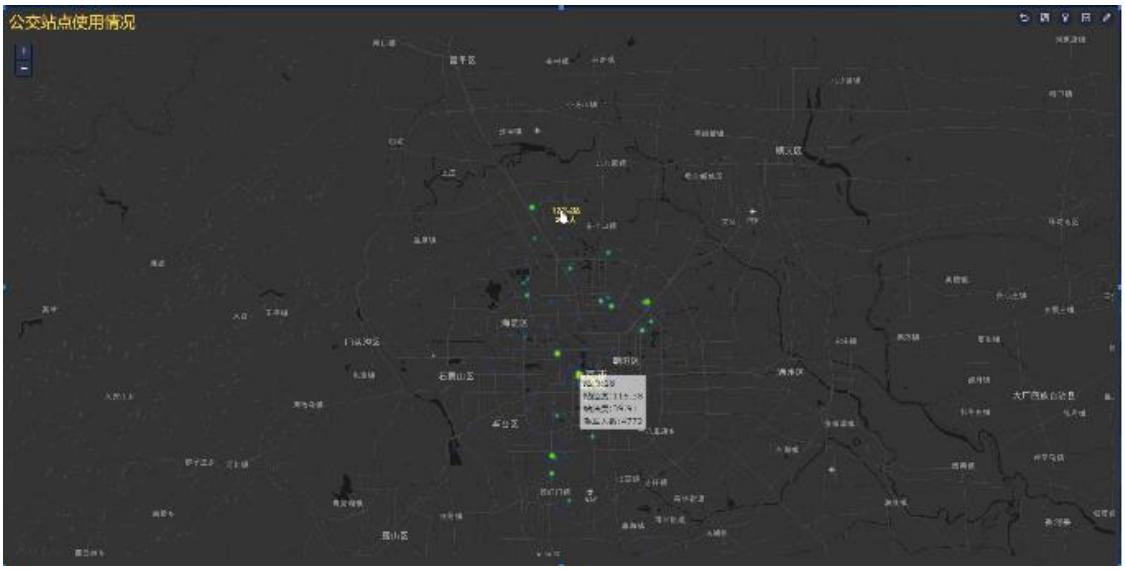
Ribecca Severino, „Dot Map“, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/dot_map.html

Heatmap (auf Karte)



Lewis Chou, „Top 10 Map Types in Data Visualization“, Medium, 26. Mai 2020, https://towardsdatascience.com/top-10-map-types-in-data-visualization-b3a80898ea70

Heat Point Map

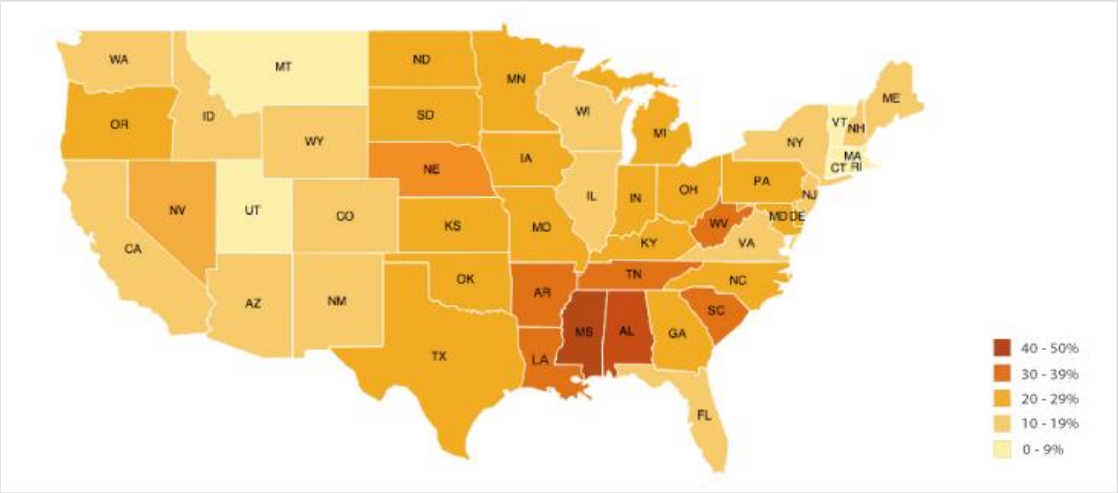


Lewis Chou, „Top 10 Map Types in Data Visualization“, Medium, 26. Mai 2020, https://towardsdatascience.com/top-10-map-types-in-data-visualization-b3a80898ea70

Connection Map

Ribecca Severino, „Connection Map “, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/connection_map.html

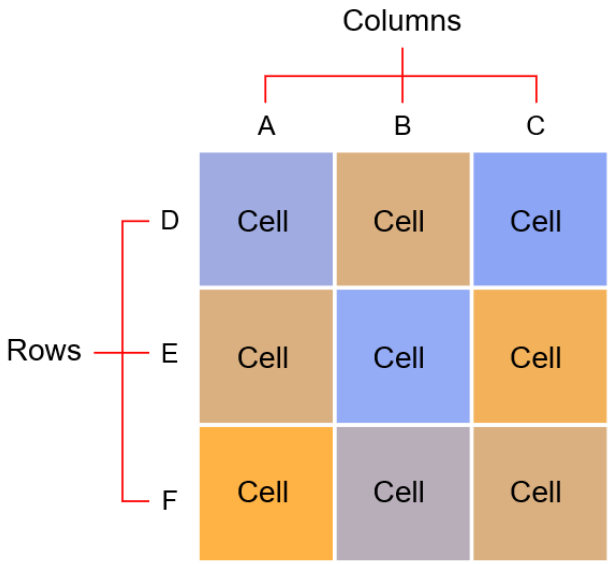
Choropleth Map



Ribecca Severino, „Choropleth Map“, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/choropleth.html

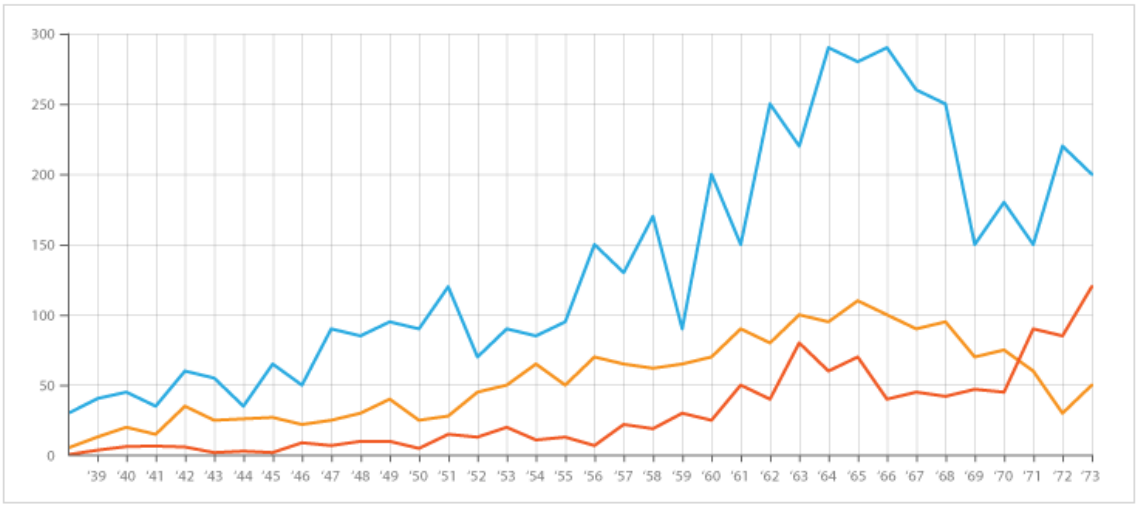
KOMBINATIONEN MIT 2D KARTEN

HEATMAP (PLUS KARTE)



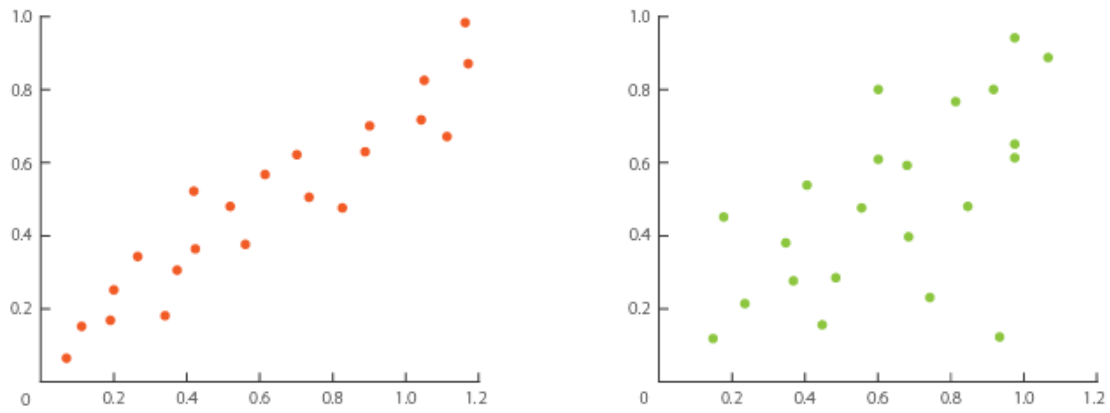
Ribecca Severino, „Heatmap (Matrix)“, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/heatmap.html

LINEGRAPH



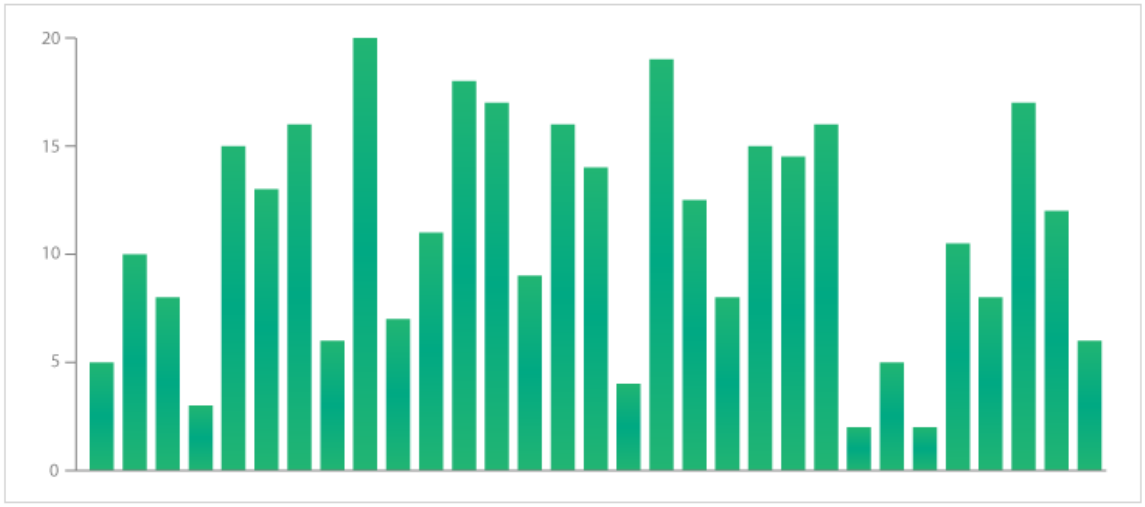
Ribecca Severino, „Line Graph“, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/line_graph.html

Scatterplot



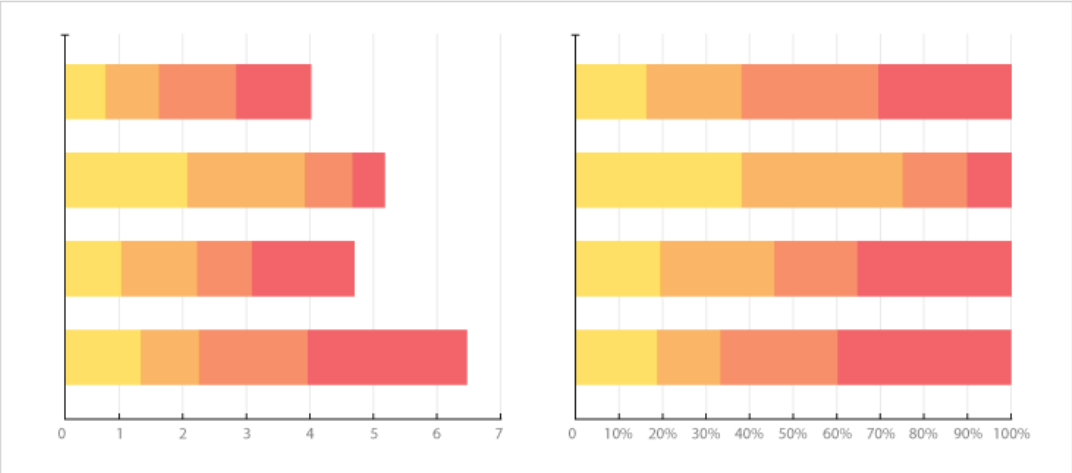
Ribecca Severino, „Scatter plot“, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/scatterplot.html

Barchart



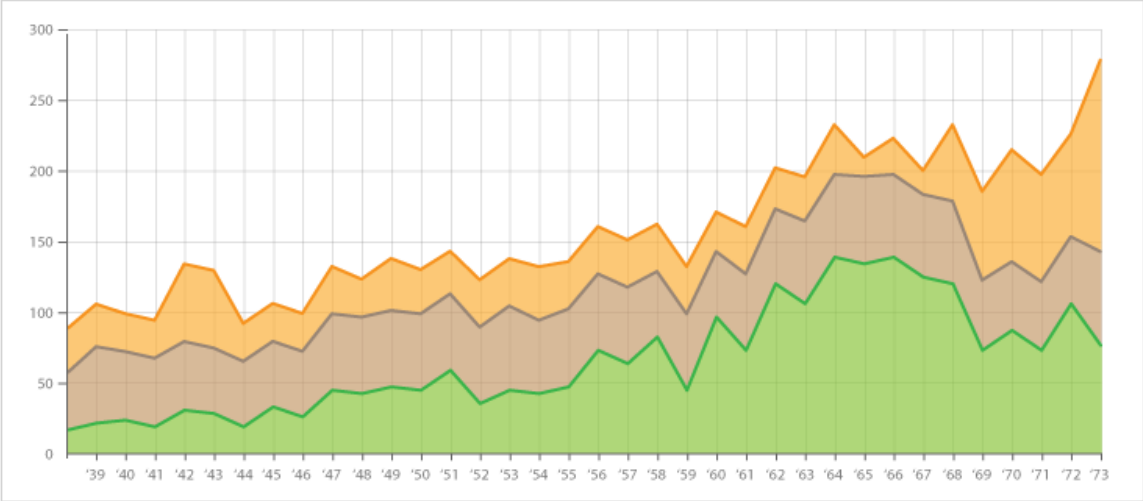
Ribecca Severino, „Bar Chart“, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/bar_chart.html

Stacked Bar Graph



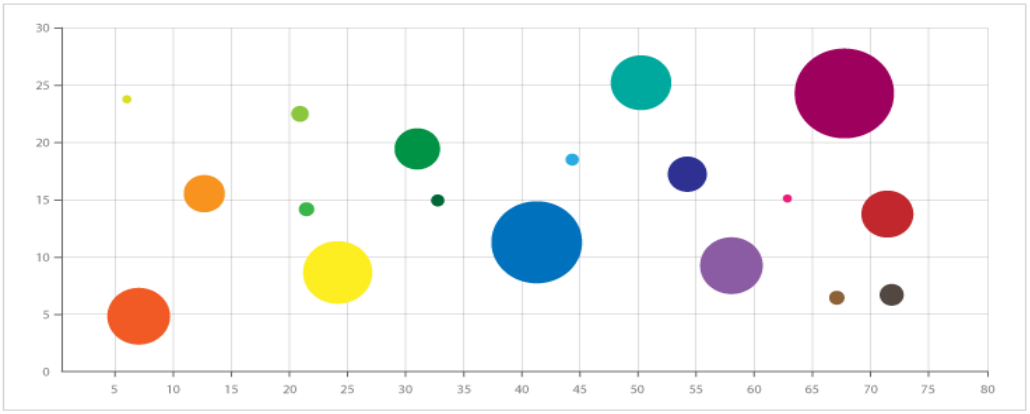
Ribecca Severino, „Stacked Bar Graph“, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/stacked_bar_graph.html

Stacked Area Graph



Ribecca Severino, „Stacked Area Graph“, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/stacked_area_graph.html

Bubble Chart



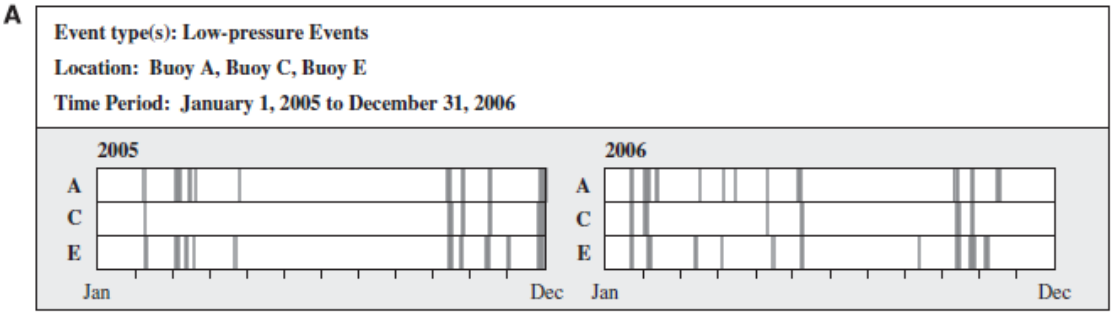
Ribecca Severino, „Bubble Chart“, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/bubble_chart.html

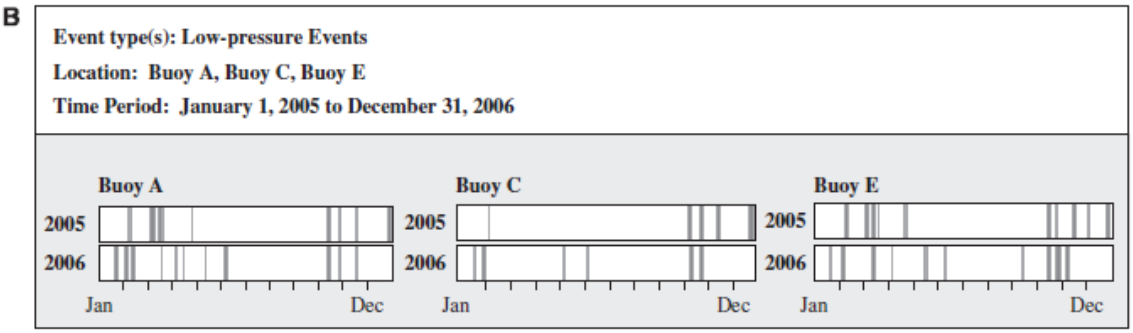
Spiral Plot

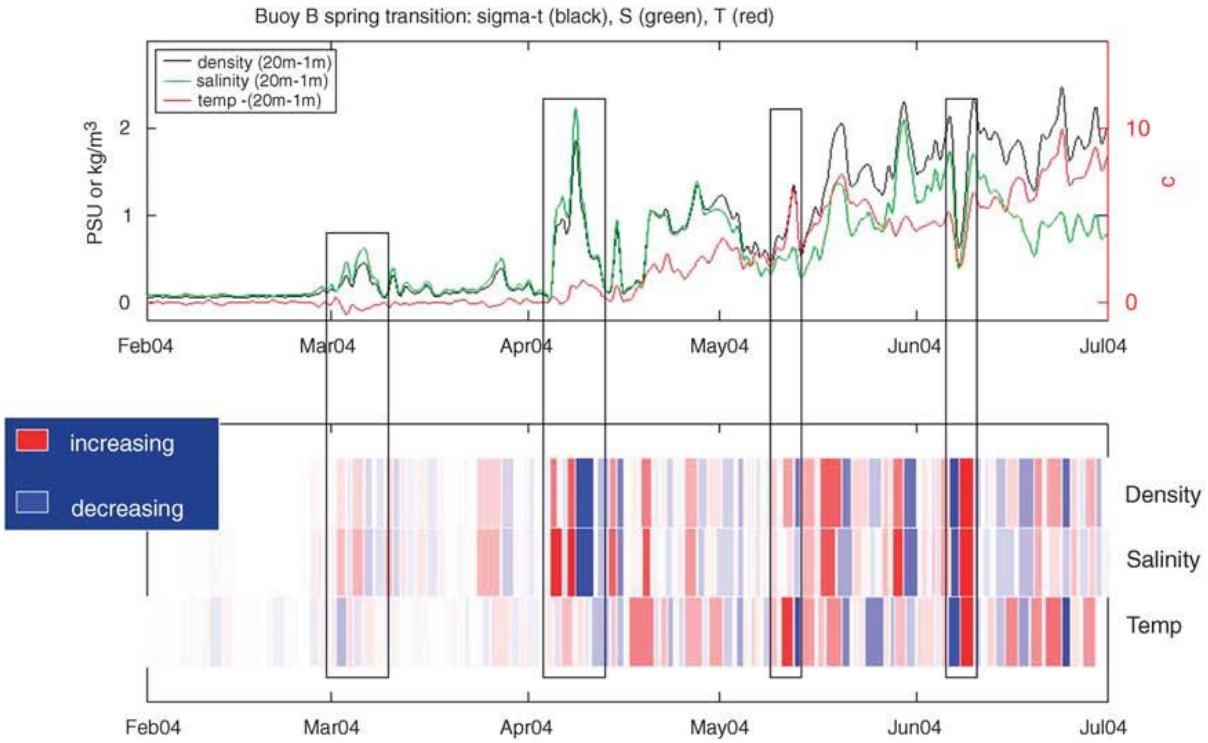
Ribecca Severino, „Spiral Plot“, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/spiral_plot.html

VISUALISIERUNG OHNE KARTEN

EVENT STACKS/EVENT VIEWER

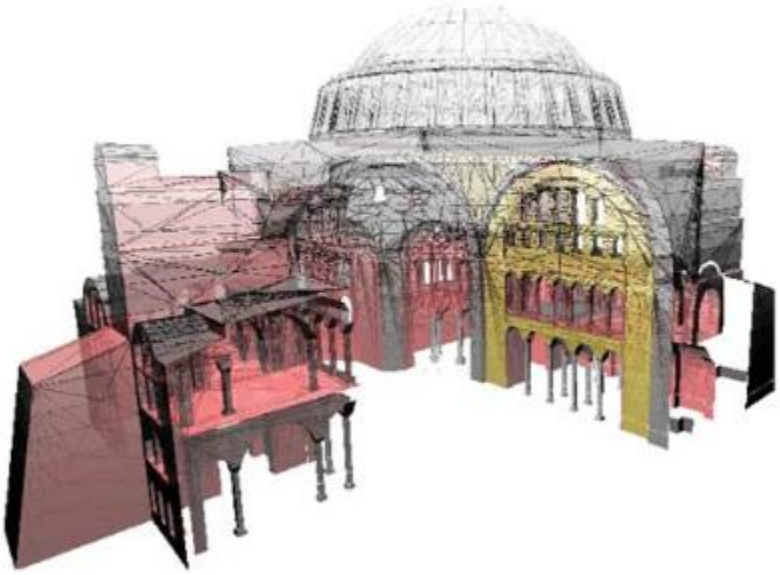






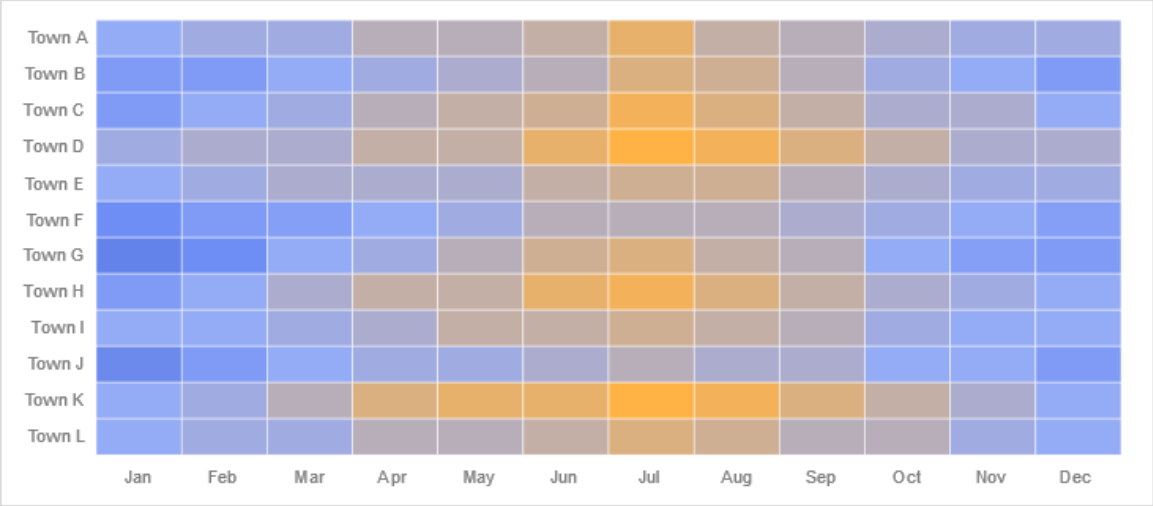
BEARD, Kate; DEESE, Heather; PETTIGREW, Neal R. A framework for visualization and exploration of events. Information Visualization, 2008, 7. Jg., Nr. 2, S. 133-151.

Temporal Focus & Context



CARVALHO, Alexandre, et al. A temporal focus+ context visualization model for handling valid-time spatial information. Information Visualization, 2008, 7. Jg., Nr. 3-4, S. 265-274.

Heatmap (ohne Karte)



Ribecca Severino, „Heatmap (Matrix)“, The Data Visualisation Catalogue, 26. Mai 2020, https://datavizcatalogue.com/methods/heatmap.html

Anhang B Icon Referenzen

- Infographic by A184 from the Noun Project
- Requirement by Nithinan Tatah from the Noun Project
- View by Sérgio Filipe Cardoso Pires from the Noun Project
- Big data by Eliricon from the Noun Project
- Radar Chart by Sam Smith from the Noun Project
- Data slide Alice Design from the Noun Project
- Abstract by Clément Payot from the Noun Project