

Visualisierung der Inlandsflüge in den USA

Data Visualisation

Abschlussaufgabe

Abgabe am 19. Mai 2023

Fakultät Wirtschaft

Studiengang Wirtschaftsinformatik

Kurs WWI2022F

von

David Kreismann

DHBW Stuttgart:

Florian Eichin

Inhaltsverzeichnis

1	Data Visualization Design Process.....	1
1.1	Einleitung.....	1
1.1	Formulating your Brief	1
1.2	Working with Data	2
1.3	Establish your editorial thinking & Develop the design solution	3
	Anhang.....	6
	Literaturverzeichnis.....	14

1 Data Visualization Design Process

1.1 Einleitung

Im Rahmen der Vorlesung Data Visualisation wurde uns von Herrn Eichin die Abschlussaufgabe gestellt, durch Analyse eines Datensatzes fünf anschauliche Visualisierungen zu erstellen. Den Datensatz können wir entweder selbst wählen oder es besteht auch die Möglichkeit den von Herrn Eichin vorgegeben Datensatz zu nutzen. Ich habe mich dabei für den vorgegeben Datensatz entschieden, welcher die Inlandsflüge aus den USA im Jahre 2015 beinhaltet. Dieser ist vom U.S. Department of Transportation (DOT) Büro für Verkehrsstatistiken publiziert und steht auf der Webseite „Kaggle“ der öffentlichen Nutzung zur Verfügung.¹

Bei der Vorgehensweise orientiere ich mich hierbei, an den in der Vorlesung vorgestellten vier Schritten des Data Visualisation Prozesses. Diese sind nach der in der Vorlesung vorgestellten Literatur „Data Visualisation: A Handbook for Data Driven Design“² aufgebaut und werden in den folgenden Kapiteln erläutert und auf mein Projekt bezogen angewendet.

Der gesamte Prozess besteht vom Grundprinzip aus zwei Hauptprozessen. Die ersten drei Schritte bilden den ersten Hauptprozess, dass sogenannte Hidden Thinking, dabei wird sich mit der Planung, Sammlung von Daten über die Erstellung der Story mit dem ganzen Planungs- und Vorbereitungsprozess beschäftigt. Dies hilft zum einen das ganze Projekt zu strukturieren. Indem es einen festen Entwicklungsprozess gibt an den sich zu halten gilt. Zum anderen lässt sich dadurch das ganze Projekt konkret planen. Der zweite Hauptprozess setzt sich dann mit der konkreten Umsetzung der Ideen aus den ersten drei Teilschritten auseinander.

1.1 Formulating your Brief

Grundsätzlich machen wir uns im folgenden Kapitel über die Rahmenbedingungen und welches Erlebnis wir dem Nutzer bieten wollen Gedanken.³

Motivation

Wie in der Einleitung schon erwähnt, ist die Motivation hinter dem Projekt die Abschlussaufgabe der Vorlesung. Die Umsetzung erfolgt somit im „professionellen“ Rahmen nach den Schritten der Vorlesung. Aus den folgenden Daten sollen möglichst fünf Visualisierungen entstehen, welchem dem Nutzer überraschende Schlüsse ziehen lassen und möglicherweise auch neues Wissen beschieren.

¹ <https://www.kaggle.com/datasets/usdot/flight-delays>

² Andy Kirk, Data Visualisation: A Handbook for Data Driven Design, 2019 2. Auflage

³ Vgl. Kirk, A. Data Visualisation: A Handbook for Data Driven Design, S. 75

Beteiligte

Die Zielgruppe / Stakeholder sind im Wesentlichen die Prüfer und Betrachter meiner Abschlussaufgabe, da diese die Aufgabe stellen und anschließend auch bewerten. Allerdings können auch weitere Personen, Kommilitonen oder Außenstehende, die Interesse an den Daten haben Teil der Zielgruppe sein. Als Visualisierer versuche ich dabei, das ganze möglichst so selbsterklärend umzusetzen, sodass auch Personen ohne wirkliches Vorwissen zu Flugzeugen und Luftfahrt ihre Schlüsse aus der visuellen Aufbereitung ziehen können.

1.2 Working with Data

Im folgenden Kapitel beschäftigen wir uns mit der Datenanalyse und wie unser Datensatz aufgebaut ist.⁴

Die verwendeten Daten

Beim verwendeten Datensatz habe ich den Originaldatensatz mir nochmal herangezogen und bearbeitet, damit dieser mit dem vorgegeben Datensatz von Herrn Eichin übereinstimmt. Dadurch erhöht sich die Anzahl der Daten und es lassen sich Prognosen für das ganze Jahr 2015 erstellen. Die Daten sind alle schon im vornherein bereinigt, es wurden nur dieselben Spalten entfernt, die auch im vorgegebenen Datensatz nicht mehr vorhanden sind. Diese beinhalteten teilweise noch die Werte zum Grund der Verspätung und die Anzahl der Minuten, die dabei verloren, gegangen sind. Der Datensatz ist ansonsten gleich aufgebaut zum vorgefertigten Datensatz und beinhaltet folgende Informationen:

- Kalenderdaten:
 - o Jahr, Monat, Tag
- Flughafendaten:
 - o Kürzel der Abflugs- und Zielflughafen
 - o Koordinaten der Abflugs- und Zielflughafen
- Flugdaten:
 - o Kürzel der Airline
 - o Flugnummer
 - o Geplante Ankunftszeit und tatsächlich benötigte Zeit (Minuten)
 - o Abflugverspätung (Minuten)
 - o Distanz (Meilen)
 - o Ankunftsverspätung (Minuten)

^{4 4} Vgl. Kirk, A. Data Visualisation: A Handbook for Data Driven Design, S. 108

1.3 Establish your editorial thinking & Develop the design solution

Im Folgenden werden die Visualisierungen beschrieben, warum ich folgende Visualisierungsart gewählt habe und was ich dem Nutzer damit zeigen möchte. Die Visualisierungen sind dabei als Dashboard umgesetzt, dies soll zum einen dem Nutzer einen größeren Überblick liefern. Indem mehrere Visualisierungen miteinander in Bezug gestellt werden können. Für das Dashboard wurde hierbei eine [Vorlage](#) genutzt⁵, diese baut auf Tailwind CSS⁶ auf und wurde im folgendem so angepasst, dass diese für folgenden Anwendungszweck genutzt werden kann. Das Dashboard besteht aus zwei Seiten. Die erste Seite mit dem Namen „Flugdaten“ soll Visualisierungen zeigen, die sich vor allem mit den Flugdaten der Airlines auseinandersetzen. Die zweite Seite mit dem Namen „Flugkarte“ soll Visualisierungen zeigen, die sich mit den Daten der Flughäfen auseinandersetzen. Durch Darstellung der Graphen auf einer Webseite, sind diese auch für Technik unerfahrene Menschen einfach einsehbar und erfordern kein großes Vorwissen zum Betrachten. Des weiteren bietet es dem Visualisierer noch viele Anpassungsmöglichkeiten seine Graphen passend darzustellen und mit Infotexten zu versorgen, so dass diese schnell verständlich sind und eine interessante Geschichte erzählen können. Anmerkung: Die Visualisierungen sind als Dashboard auch online verfügbar unter der URL: davidkreismann.me. Zu jeder Visualisierung gibt es auch einen kurzen Infotext. Dieser lässt sich öffnen durch Drücken des Info-Icons neben der Überschrift zu jeder Visualisierung.

Der Betrachter soll aus den Visualisierungen spannende Informationen ziehen können, aber auch durch Daten in Relation setzen seine eigene Meinung bilden können. Dabei kann er sich folgende Thesen bspw. stellen: Welche Airline hat die wenigste Verspätung? Welche Airlines sind die größten in den USA? Gibt es Korrelationen zwischen Verspätungszeit und Größe der Airline? Welche Flughäfen haben die meisten Verspätungen? Wie steht das in Zusammenhang mit der Größe des Flughafens? Welche Airline fliegt welche Flughäfen an?

Die schnellste Airline / Scatterplot

Die erste Visualisierung, siehe Abbildung 2, soll einen Überblick darüber geben, welche Airline im Durchschnitt pro Monat die wenigste bzw. die meiste Verspätungszeit bei Ankunft hatte. Das Ganze wird deshalb in einem Scatterplot dargestellt, wobei alle Punkte einer Airline mit einer Linie verbunden werden. Auf der X-Achse soll dabei die Verspätungszeit in Minuten zu sehen sein und auf der Y-Achse die Monate von Januar bis Dezember. Der Übersichtlichkeit halber ändert sich die Farbe der Linie von Rot zu Grün, je nachdem wie hoch die Verspätungszeit in dem Monat ist. Auf Grund der Anzahl der Airlines überschneiden sich einige der Linien stark miteinander, deshalb soll ein Hover-Effekt eingebaut werden, wenn der Nutzer über die

⁵ <https://www.tailwindtoolbox.com/templates/minimal-admin-template>

⁶ <https://tailwindui.com/>

einzelnen Punkte fährt. Auf den Dashboxen sollen dann auch noch Informationen zur Airline und Verspätungszeit entnommen werden können, beim drüber schweben über die Punkte.

Aus den Daten lässt sich somit erkennen, welche Airline, die wenigste Verspätungszeit hatte oder welche Airline in einem Monat schlecht performt hatte. Es lassen sich also verschiedene Airlines miteinander im Vergleich setzen. Wir können jedoch aus den Durchschnittsdaten keinen Entschluss ziehen, ob die Verspätung durch einen großen Ausfall zu Stande gekommen ist oder ob die Airline viele kleinere Verspätungen in dem Monat hatte.

Anzahl der Flüge pro Airline und Distanz in 1000 Meilen / Bar Chart

Die zweite Visualisierung, siehe Abbildung 4, soll für jede Airline die Anzahl der Flüge und die geflogene Distanz über das Jahr 2015 zeigen. Für solche festen Werte bietet sich ein Bar Chart an. Auf der X-Achse sollen die Werte zu sehen sein und auf der Y-Achse die ganzen Airlines gelistet, da es sich hierbei um zwei Werte pro Airline handelt muss auf eine kontrastreiche Farbkombination geachtet werden, damit die einzelnen Balken der Anzahl der Flüge bzw. der Distanz zugeordnet werden können.

Aus den Daten lässt sich zum einen erkennen, wie groß eine Airline voraussichtlich ist, da für mehr Flüge auch mehr Flugzeuge und Personal benötigt werden. Darüber hinaus lässt sich erkennen, dass vor allem die großen Airlines im Durchschnitt mehr Flüge durchführen in Korrelation zu der Distanz. Während bei den kleinen Airlines die Anzahl der Flüge mit der Distanz eher gleich auf ist.

Verzögerungszeit der Airlines / Boxplot

Die dritte Visualisierung, siehe Abbildung 5, soll die Verzögerungszeit für jede Airline für den 01. Januar 2015 zeigen, dabei werden jedoch nur Werte größer 0 berücksichtigt. Da es auch vorkommen kann, dass ein Flug vor der geplanten Zeit ankommt. In dem Fall eignet sich ein Boxplot, da aus diesen im Gegensatz zum Balkendiagramm noch viele weitere Informationen gezogen werden können. Auf der X-Achse haben wir die Verspätungszeit und auf der Y-Achse wieder die ganzen Airlines. Die X-Achse ist nur bis zum Wert 100 gesetzt, um die Boxplots übersichtlicher erscheinen zu lassen. Wie sich jedoch erkennen lässt gibt es vereinzelt Airlines die viel stärkere Ausreiser haben. Diese lassen sich jedoch beim drüber schweben mit der Maus über die Boxplots ablesen.

Mit diesen Daten lässt sich somit zum einen ein detailliertes Bild ablesen, welches auch nochmal im Vergleich mit der ersten Visualisierung genauere Informationen liefert für den Monat Januar. Es könnte zum Beispiel abgeleitet werden, dass bei der Anzahl an Flügen einzelne Verspätungen keinen zu großen Effekt haben auf die Verspätungszeit.

Anzahl der Flüge von den Flughäfen / Treemap

Die vierte Visualisierung, siehe Abbildung 10, soll die 30 größten Ankunftsflughäfen nach Anzahl, sowie die 30 größten Abflughäfen anzeigen. In dem Fall habe ich mich für eine Treemap entschieden, da diese sehr übersichtlich und schnell erkennbar durch ihre Größe der Rechtecke die Werte in Relation setzt. Dabei sind die Flughäfen mit ihrem Kürzel abgebildet und die absolute Anzahl der Flüge ist unter dem Kürzel zu finden.

Aus diesen Daten konnte ich ablesen, dass jeder Flughafen fast dieselbe Anzahl an Abflügen und Ankünften hat über das Jahr verteilt, das ist nach kurzem Überlegen natürlich auch nicht anders möglich, da Flugzeuge natürlich wieder abheben müssen, damit neue landen können. Das ganze jedoch in konkreten Zahlen abgebildet zu sehen, finde ich dennoch interessant. Die Flughäfen belegen somit eins zu eins dieselben Plätze in beiden Kategorien.

Verspätungszeit der Flughäfen & Flugrouten / Bubble Maps & Lines on Maps

Die fünfte und letzte Visualisierung, die in Abbildung 7 dargestellt wird, soll eine Karte der USA zeigen, die nach Verspätungszeit Punkte an den jeweiligen Koordinaten der Flughäfen zeigt. Ich habe mich für eine Bubble Map entschieden, da wir die Punkte mit einem bestimmten Wert versehen können. Je nach Verzögerungszeit können diese Punkte dann größer oder kleiner sein. Hierbei ist zu beachten, dass es einzelne Flughäfen mit dem Wert: „0,1“ gibt. Diesen Wert musste ich einigen Flughäfen zuteilen, damit alle Flughäfen abgebildet werden können, denn diese Flughäfen hatten eigentlich im Durchschnitt eine negative Ankunftszeit, also Flugzeuge sind eher gelandet als geplant. Das Ganze lässt sich dann über ein Dropdown-Menü nach Abflugs- und Ankunftszeit Verspätung filtern. Zudem gibt es ein weiteres Dropdown-Menü. Indem nach den entsprechenden Flugrouten der Airlines vom 01. Januar 2015 gefiltert werden kann. Bei dieser Filterung werden dir dann die entsprechenden Flugrouten der Airline per Linien gekennzeichnet inklusive Abflugs- und Ankunftsflughafen, welche mit einem roten Punkt gekennzeichnet sind und beim drüber schweben mit der Maus abgelesen werden können. Dafür habe ich eine Lines on Maps Darstellung verwendet.

Aus dieser Visualisierung lassen sich mehrere Erkenntnisse ziehen. Zwei wichtige Erkenntnisse, sind zum einen, wenn die vierte Visualisierung mit dieser Visualisierung in Verhältnis gesetzt wird, lässt sich überraschend erkennen, dass die Flughäfen mit der meisten Verspätung nicht in Zusammenhang mit der Anzahl der Flüge pro Flughafen in Zusammenhang stehen. Die zweite Erkenntnis, die mich überrascht hatte, dass bei den Flugrouten viele Airlines sich auf bestimmte Gebiete teilweise spezialisiert haben. Bestimmte Airlines fliegen bspw. nur den Westen der USA bzw. den Osten der USA an. Zudem haben einige Airlines bestimmte Flughäfen, welchen sie als Drehpunkt benutzen. Von wo aus, dann die meisten Flüge abfliegen.

Anhang

Anhang 1: Flugdaten Dashboard – Flugdaten	7
Anhang 2: Die schnellste Airline	7
Anhang 3: Die schnellste Airline – mit Hover-Effekt	8
Anhang 4: Anzahl der Flüge pro Airline und Distanz in 1000 Meilen.....	8
Anhang 5: Verzögerungszeit der Airlines (in Minuten)	9
Anhang 6: Flugdaten Dashboard - Flugkarte.....	9
Anhang 7: Flugkarte.....	10
Anhang 8: Flugkarte mit Flughäfen Filterung.....	10
Anhang 9: Flugkarte mit Flugroutenfilterung.....	11
Anhang 10: Anzahl der Flüge von den Flughäfen	12
Anhang 11: Flugdaten Dashboard - Infofeld	13
Anhang 12: Flugdaten Dashboard – Sidebar zur Dokumentation	13

Anhang 1: Flugdaten Dashboard – Flugdaten

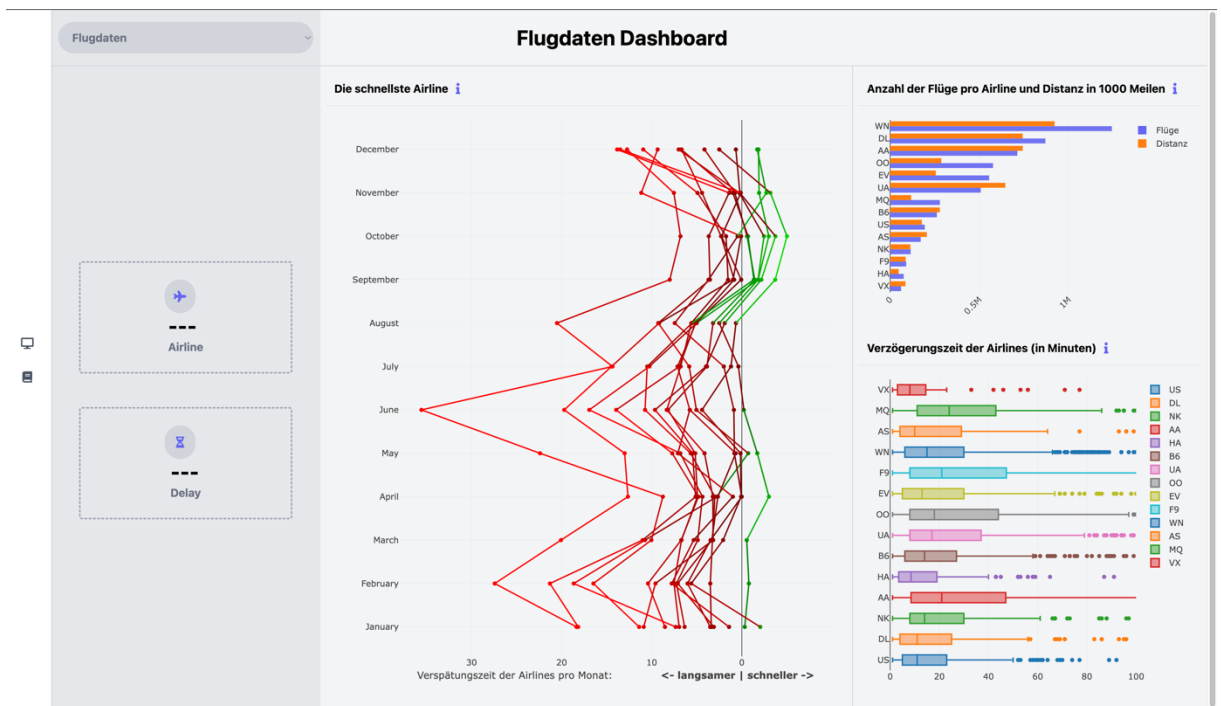


Abbildung 1: Flugdaten Dashboard – Flugdaten

Anhang 2: Die schnellste Airline

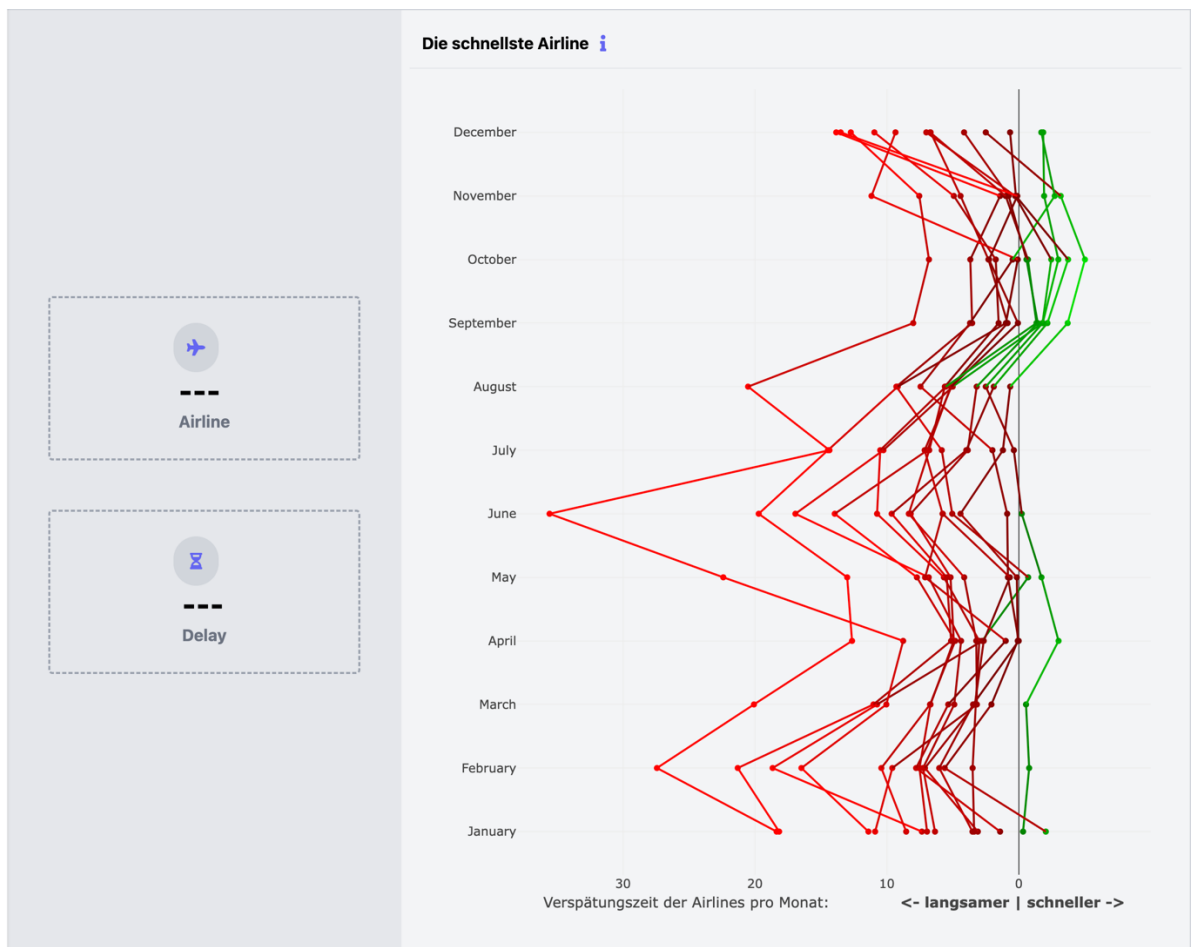


Abbildung 2: Die schnellste Airline

Anhang 3: Die schnellste Airline – mit Hover-Effekt

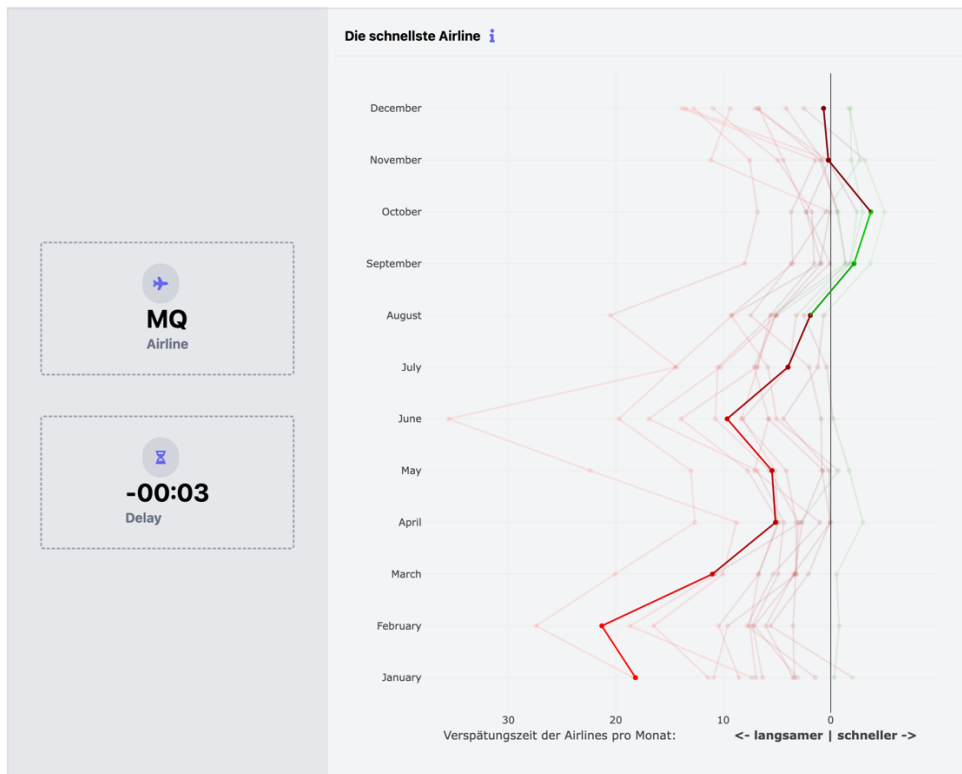


Abbildung 3: Die schnellste Airline – mit Hover-Effekt

Anhang 4: Anzahl der Flüge pro Airline und Distanz in 1000 Meilen

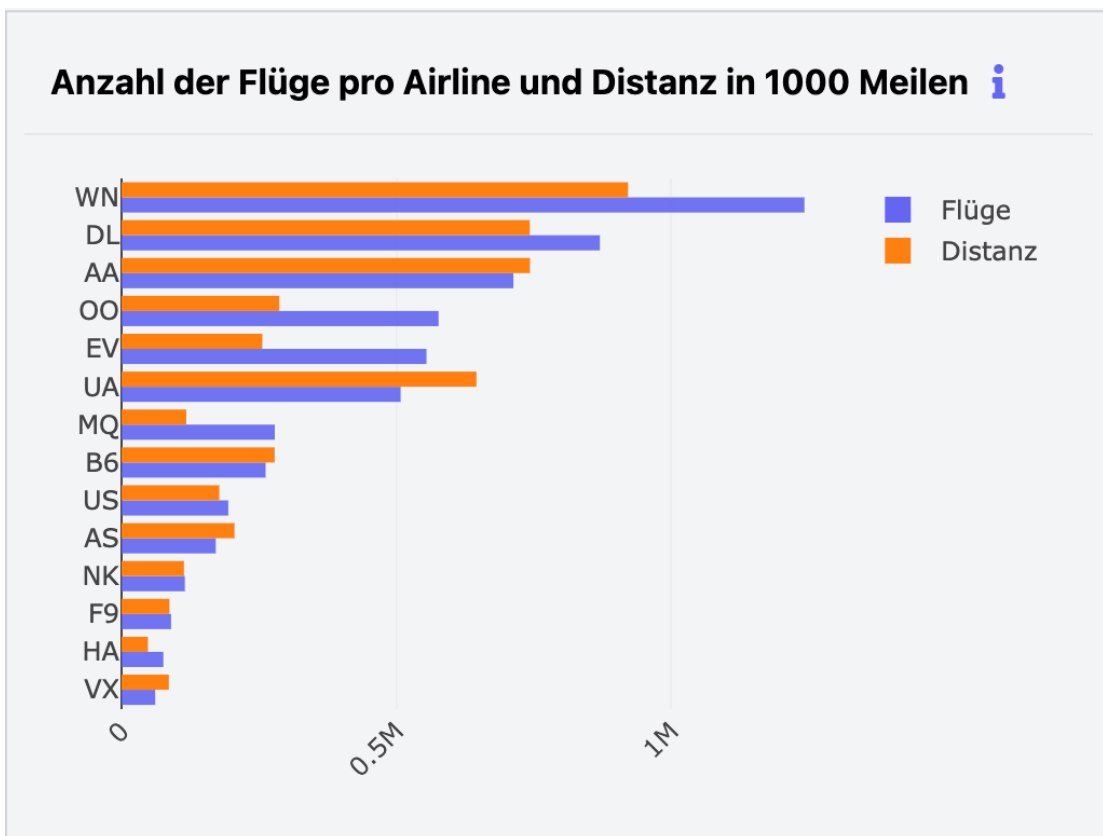


Abbildung 4: Anzahl der Flüge pro Airline und Distanz in 1000 Meilen

Anhang 5: Verzögerungszeit der Airlines (in Minuten)

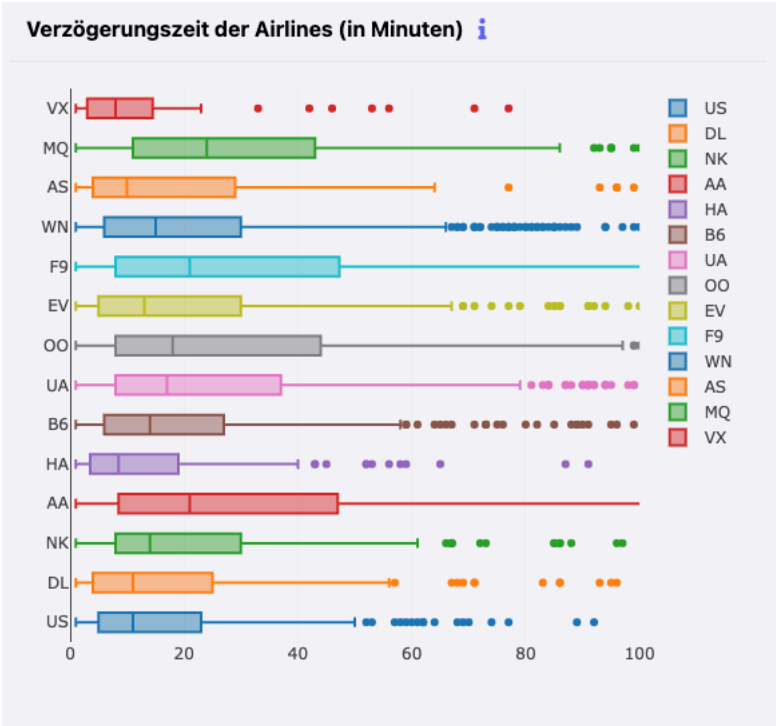


Abbildung 5: Verzögerungszeit der Airlines (in Minuten)

Anhang 6: Flugdaten Dashboard - Flugkarte

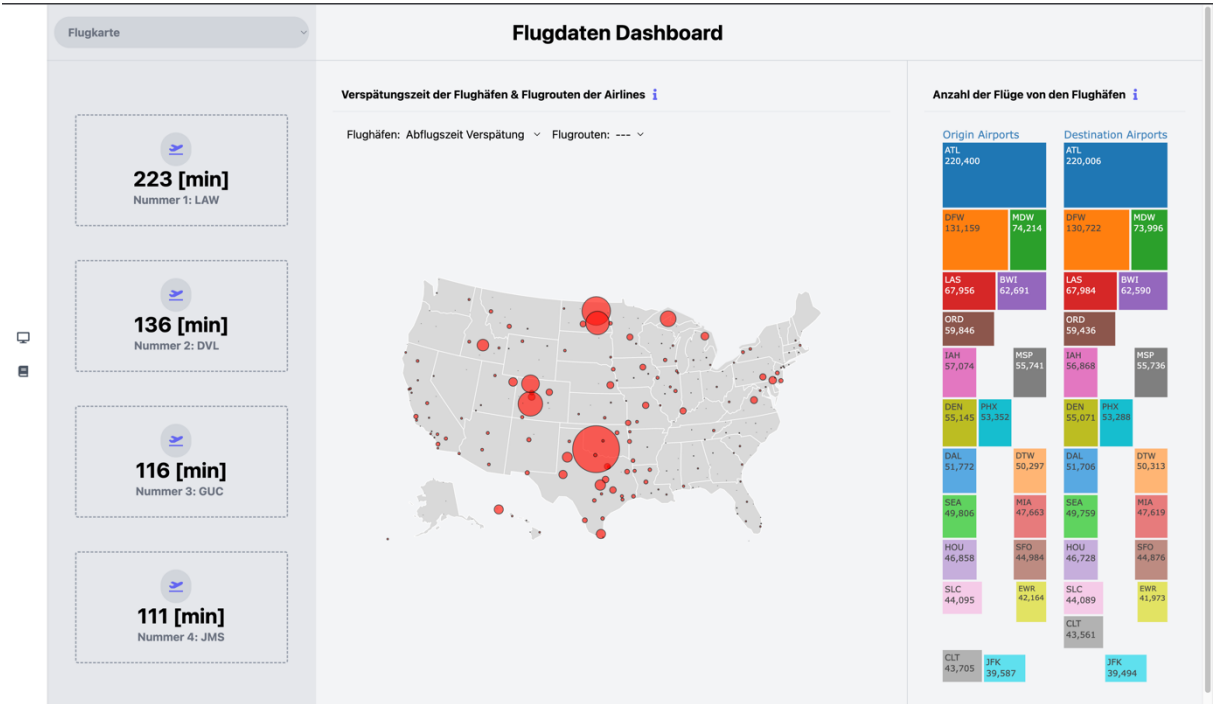


Abbildung 6: Flugdaten Dashboard - Flugkarte

Anhang 7: Flugkarte

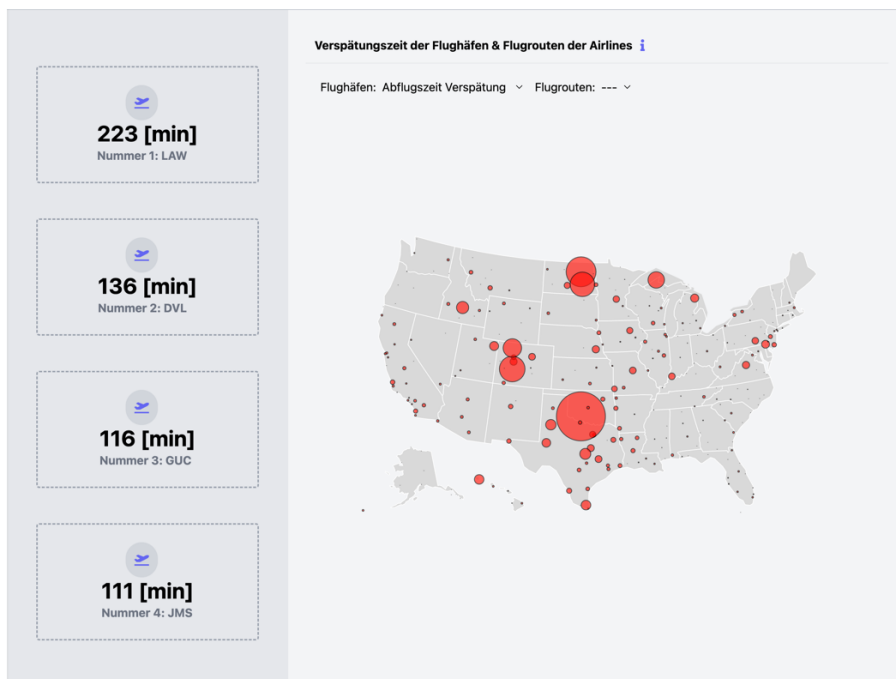


Abbildung 7: Flugkarte

Anhang 8: Flugkarte mit Flughäfen Filterung

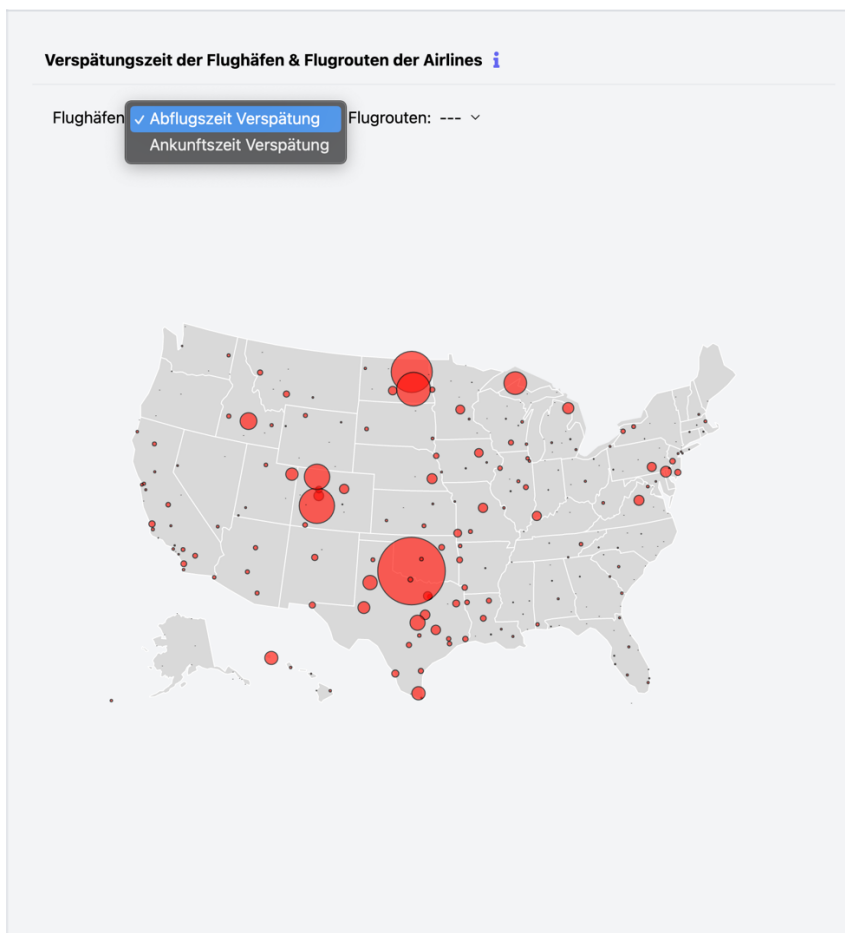


Abbildung 8: Flugkarte mit Flughäfen Filterung

Anhang 9: Flugkarte mit Flugroutenfilterung



Abbildung 9: Flugkarte mit Flugroutenfilterung

Anhang 10: Anzahl der Flüge von den Flughäfen

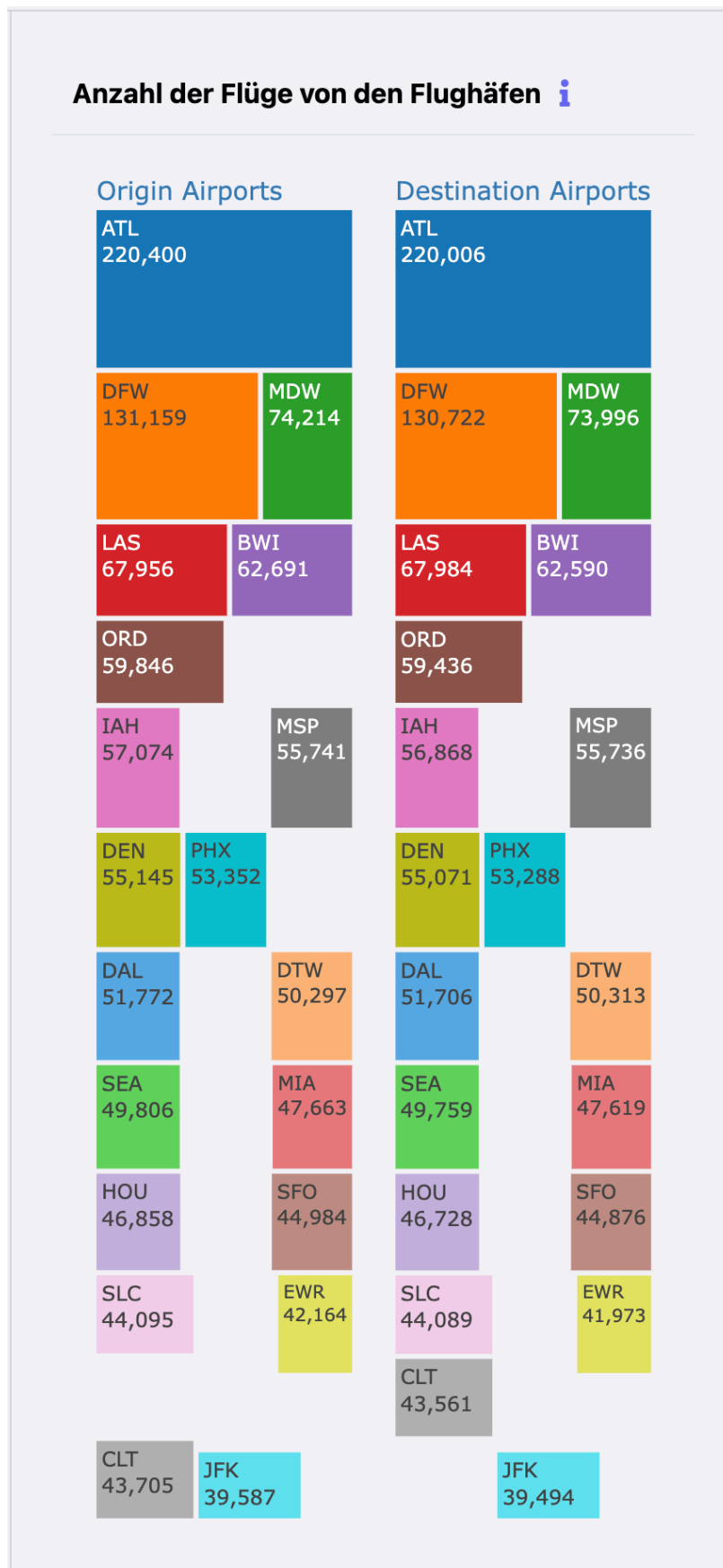


Abbildung 10: Anzahl der Flüge von den Flughäfen

Anhang 11: Flugdaten Dashboard - Infocfeld

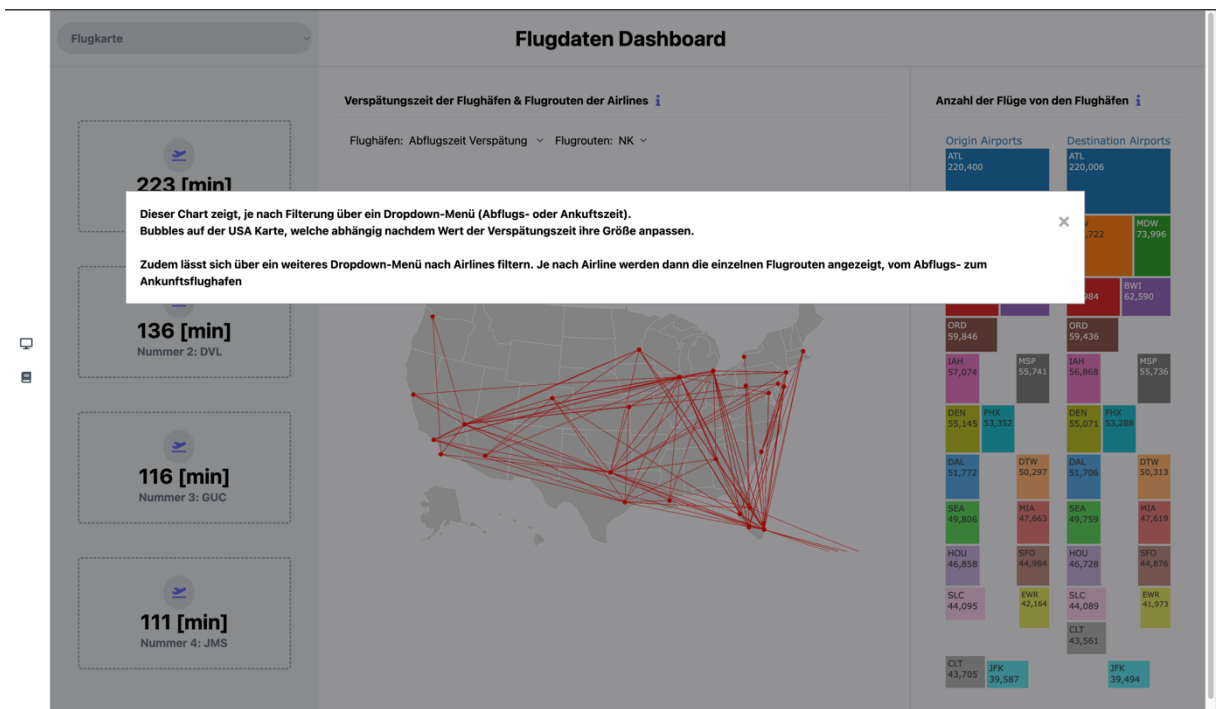


Abbildung 11: Flugdaten Dashboard - Infocfeld

Anhang 12: Flugdaten Dashboard – Sidebar zur Dokumentation

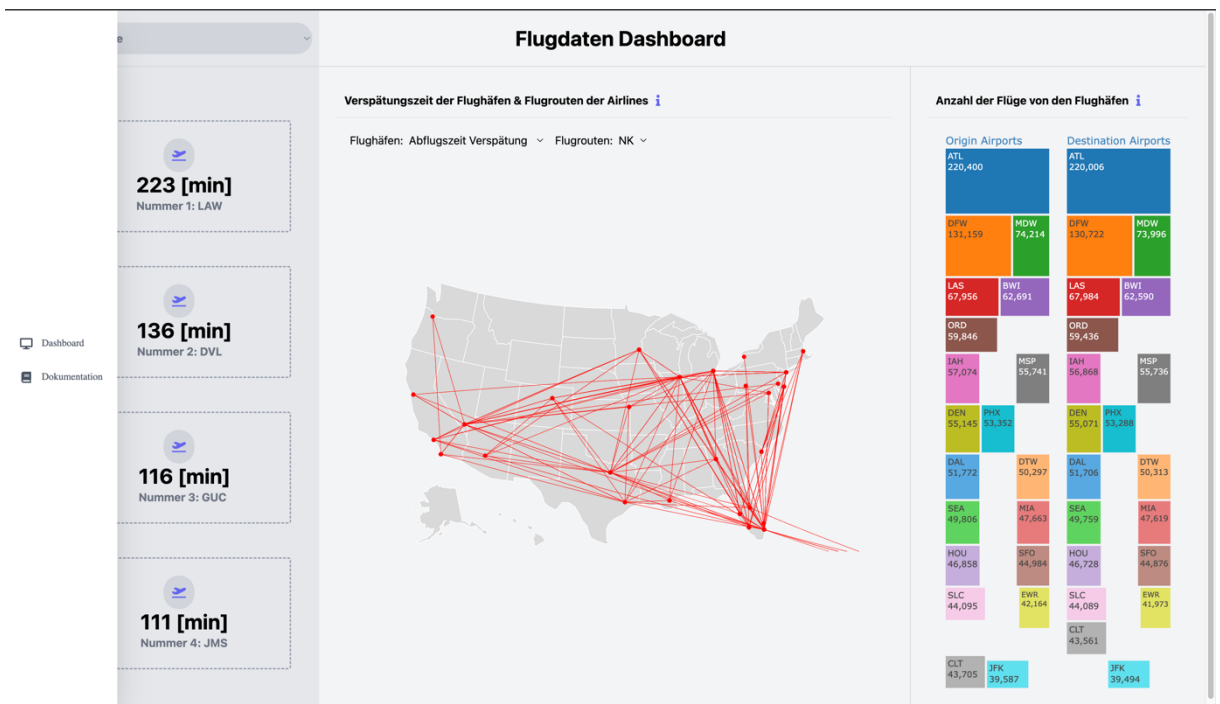


Abbildung 12: Flugdaten Dashboard – Sidebar zur Dokumentation

Literaturverzeichnis

Kirk, A. (2019): Data Visualisation: A Handbook for Data Driven Design, 2. Auflage

Erklärung

Ich versichere hiermit, dass ich meine *Abschlussaufgabe mit dem Thema: Data Visualisation* selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich versichere zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

(Ort, Datum)

(Unterschrift)