

# **Dokumentation der Abschluss- aufgabe in Data- Science Visualization**

---

**Von Tobias Hölzer**

---

## Abstract

Im Laufe des Kurses Data Science Visualization wurde dem Kurs die Aufgabe gegeben drei bis fünf Grafiken zu erstellen, um mögliche Fragestellungen bezüglich des Flugverkehrs in Amerika am 01. Januar 2015 zu beantworten. Bei dieser Aufgabe geht es primär um die Visualisierungen, weniger um die Data Science dahinter. Die Flugdaten für den 01. Januar 2015 aus Amerika bekamen wir bereitgestellt, diese beinhalteten Daten wie Abflugzeit, Verspätung etc. zu jedem einzelnen Flug, der an diesem Tag losgeflogen ist.

## Zielsetzung

Ich wollte für meine Visualisierungen, dass der Benutzer wählen kann, wie er sein visualisiertes Daten Set zusammengestellt haben möchte. Meine Visualisierungen sind daher alle bis auf eins durch Wahlmöglichkeiten unterstützt. Meine Visualisierungen sind eine interaktive Scatter-Map, ein interaktives Bar Chart und ein Pie Chart, das Pie Chart ist das einzige, welches nicht durch Wahlmöglichkeiten unterstützt wird. Alle sind Browsergestützt.

Das Ziel hinter den Visualisierungen sollte sein, dass der Benutzer anhand der Visualisierung seine eigenen Schlüsse bezüglich der Verspätungen von einzelnen Flügen, Airlines und Flughäfen ziehen kann. Daher habe ich so viele mir zur Verfügung stehenden Verspätungsangaben in die Benutzerauswahl mit eingebracht.

## Stakeholder

Ein typischer Benutzer/ Stakeholder wäre dementsprechend ein Fluggast, welcher Probleme mit Verspätungen hat, ein Journalist, welcher über Verspätungen im Flugverkehr recherchieren möchte, aber auch einzelne Flughäfen und Airlines, welche Ihre Verspätungen reduzieren möchten.

## Art der Visualisierung

Durch eine Scatter-Map können sehr gut die Koordinaten des Flughafens, sowie durch Farbe und Größe des Punktes, welcher den Flughafen repräsentiert weitere Eigenschaften eines Flughafens gezeigt werden. Somit kann man auf einer anschaulichen zweidimensionalen Fläche dreidimensionale Daten anzeigen.

Da die Scatter-Map jedoch zwar Koordinaten genau anzeigen kann, jedoch eher weniger die durch Größe und Farbe repräsentierten absoluten Werte. Ein Bar Chart hingegen kann diese Absoluten Werte sehr gut anzeigen, jedoch aufgrund seiner zweidimensionalen Darstellungsweise nur einen Wert bezüglich eines Flughafens.

Um die zahlentechnische Beziehung zwischen den besten Werten und dem durchschnittlichen Rest darzustellen eignet sich besonders ein Pie Chart, da dieses Zahlengrößen sehr gut darstellen kann.

## Design

Das Design sollte schlicht, responsive und UX-basiert sein, also einfach für den Benutzer zu benutzen und verstehen sein, auf allen Geräten und vollem Fokus auf die dargestellten Daten bieten. Als Form der Aufgabenlösung wählte ich zusammen mit Philipp aus meinem Kurs die Form eines Dashboards, wo alle Darstellungen zusammengefasst und an einem Ort stehen. Seinen Teil findet man im „Wetter“-Tab, meinem im „Verspätungen“-Tab. Um meinen Text und die Selectors schön und clean darzustellen benutze ich Googles Material Web Components. Außerdem wählte ich einen beruhigenden, dunklen blau-grauen Farbton als primäre Farbe für das Dashboard. Außerdem sind alle Grafiken durch Plotly unterstützt, weshalb dem Benutzer weitere Auswahlmöglichkeiten bezüglich der Anzeige, sowie ein Download in Form eines Bildes zur Verfügung stehen.

---

## Die Map

Das primäre Ziel der Map (Map Scatter Box) war es, den verschiedenen Flughäfen Verspätungszeiten zuzuweisen. Hier soll der Benutzer schnell Ausreißer erkennen können. Außerdem sollten noch zwei weitere Komponenten mit eingebracht werden, die Flugzeit und die Aufgeholte Verspätung.

## Datenlage und Operatorauswahl

Die Datenlage sieht wie folgt aus: Jeder Flughafen empfängt und entsendet mehrere Flugzeuge. Das grundlegende Daten Set beherbergt jedoch nur die Daten jeden einzelnen Fluges. Daher wurden daraus die Daten für die einzelnen Flughäfen interpretiert. Um zum Beispiel die Verspätung zur Ankunft am Flughafen zu berechnen, können die Verspätungen der einzelnen Flüge, welche an dem Flughafen gelandet sind, entweder zusammen berechnet werden oder aus allen Verspätungen der Durchschnitt ermittelt werden. Der Benutzer kann unter der Operator Sektion wählen, ob er die Daten als Absolut Werte oder als Durchschnittswerte berechnet haben möchte.

## Koordinaten und Flughafenauswahl

Die einzelnen Punkte an den jeweiligen Koordinaten zeigen die Standorte der verschiedenen Flughäfen an. Auf der rechten Seite hat der Benutzer die Möglichkeit zwischen Abflug Flughafen und Ankunft Flughafen wählen. Dies bedeutet, dass wenn der Benutzer zum Beispiel Abflugs Flughafen ausgewählt hat, dass zum einen nur Flughäfen angezeigt werden, an denen an diesem Tag auch Flugzeuge abgeflogen sind und zum anderen, dass alle Daten bezüglich des Flughafens von Flügen

kommen, welche an diesem Flughafen abgeflogen sind. Mit Daten bezüglich des Flughafens sind die durch Größe und Farbe repräsentierte Verspätung und Flugzeit/ Aufgeholte Verspätung gemeint.

## Größe der Punkte und Verspätungsauswahl

Die Größe der Punkte repräsentiert die Verspätung der Flüge, je größer ein Punkt auf der Karte, umso größer die Verspätung. Um welche Verspätung es sich handelt ist dem Benutzer überlassen, dieser kann wählen zwischen Verspätung bei Abflug und Verspätung bei Ankunft wählen. Gepaart mit den Wahlmöglichkeiten des Flughafens ist es nun also zum Beispiel möglich zu sehen, welche Flughäfen empfangen Flüge mit besonders hoher Verspätung beim Start an einem anderen Flughafen. Diese zwei Kausalitäten scheinen auf den ersten Blick recht wenig miteinander zu tun zu haben, jedoch ist auf der Karte ersichtlich, dass zum Beispiel der Flughafen JMS im Schnitt besonders viel Pech mit seinen empfangenen Flügen hat.

## Farbe der Punkte und Auswahl der 3. Komponente (Zeit)

Neben Verspätung und Flughafen kann der Benutzer auch noch eine 3. Komponente wählen, zusammengefasst unter das Wort Zeit. Die Farbe der einzelnen Punkte repräsentiert die Daten dieser ausgewählten 3. Komponente, auf der rechten Seite der Map befindet sich eine Skala dazu. Der Benutzer kann an dieser Stelle wählen, ob die Farbe die Flugzeit der Flüge oder die Aufgeholte Verspätung je 1000km anzeigen soll. Verspätung je 1000km daher, um Vergleichbarkeit herzustellen: Einem langen Flug wird es leichter fallen Verspätung einzuholen als einem kurzen Flug.

## Ausblick

Rein theoretisch könnte man die Anzahl an Auswahlmöglichkeiten noch weiter vergrößern, indem der Operator getrennt für die Verspätungsauswahl und die Auswahl der 3. Komponente wählbar ist. Ebenso könnte man weitere Auswahlmöglichkeiten der 3. Komponente aber auch der Verspätungskomponente hinzufügen, wie zum Beispiel die Verspätung bei Landung vs. Die Verspätung bei Ankunft. (Flugzeug setzt am Boden auf vs. Passagiere können aussteigen)

---

## Der Bar Plot

An sich ist das Bar Plot ähnlich der Map aufgebaut, da Ihr Zweck die Darstellung von konkreten Daten aus der Map ist. Jedoch besitzt das Bar Chart nur 2 Dimensionen, weshalb die Verspätungs-

Komponente und die 3. Komponente hier in einer Komponente vertreten sind. Auch hier kann der Benutzer den Rechenoperator wählen.

## X-Achse

Die X-Achse ist das Äquivalent zu der Flughafen-Komponente. Jedoch kann der Nutzer hier neben Ankunft- und Abflugflughafen auch noch zwischen Airline und Flugnummer wählen.

## Y-Achse

Hier kann der Benutzer zwischen Verspätung bei Ankunft, Verspätung bei Abflug, Flugzeit und Aufgeholte Verspätung.

---

## Das Pie Chart

Hier werden die 5 Flugnummern angezeigt, welche durchschnittlich die größte Verspätung haben. Bei hover wird die Verspätung in Minuten angezeigt. Dazu wird zum Vergleich die durchschnittliche Verspätung aller anderen Flugnummern im Durchschnitt angezeigt. So können anschaulich die Flüge mit den größten Verspätungen angezeigt werden und inwiefern sich diese von der „normalen“ Verspätungszeit unterscheidet. Die Werte onhover sind korrekt, jedoch sind die Prozentzahlen mit Vorsicht zu genießen: Hier wird die Verspätung der Restlichen Flüge nochmals als Durchschnitt genommen, weshalb die Prozente sich nun nicht mehr den Anteil aller durchschnittlichen Verspätungen angibt, sondern immer den Anteil an der Gesamtheit der Angezeigten werden. Blendet man zum Beispiel den Rest durch klicken auf die Legende aus, so zeigt der Prozentsatz den Anteil an Verspätungen unter den 5 meistverspäteten Flüge.