



puenktlichkeit.ch

Analysen zum öV Schweiz

Dokumentation Autorenprojekt

Studiengang: CAS Datenvisualisierung, Sommer 2017

Autor: Andreas Gutweniger

Datum: 29. September 2017

Abstract

Etwas transportieren. Züge transportieren Passagiere. Infografiken transportieren eine Aussage. Züge fahren manchmal leer durch die Landschaft. Und Infografiken?

Mein Datensatz, das sind Pünktlichkeitsdaten des öffentlichen Verkehrs in der Schweiz. Von 28 Unternehmen. Mit 2'300 Haltestellen. An bereits 290 Tagen. Jeden Tag 260'000 Halte. Mittlerweile sind es 75 Millionen. Das ist schon ein wenig «Big Data». Viel Material für Datenanalyse. Viel Material für Machine Learning. Viel Material für Story Telling?

Ganz ehrlich: Verspätungen passieren aus den unterschiedlichsten Gründen. Hier eine defekte Weiche, dort ein Personunfall, hier eine Schulreise, dort ein Blitzeinschlag. Das meiste davon völlig unsystematisch. Und so ist der Grossteil meiner Daten das, was Physiker als «Rauschen» bezeichnen: absolut zufällig – und absolut langweilig.

Und das ist gut so. Denn es zeugt vom hohen Niveau des öV Schweiz: Systematische Störungen sind weitgehend ausgemerzt. Weitgehend, aber nicht vollständig.

Mit meinem Projekt puenktlichkeit.ch soll der Benutzer entdecken können, wo es Muster in den Daten gibt. Weil es ihn betrifft (als Fahrgast). Oder weil er den öV noch besser machen will (als Mitarbeiter). Jeden Tag neu. Denn jeden Tag gibt es neue Daten. Und jeden Tag ein neues Bild auf puenktlichkeit.ch.

Wir verstehen am besten, was wir selbst entdeckt haben. Entdecken kann man nur, wenn man sucht. Vielleicht ist dies der nobelste Beitrag von Datenvisualisierung: Die Suche nach «Patterns» brauchen wir nicht den Algorithmen zu überlassen.

Und wenn es keine Story gibt? «Publication Bias» ist ein gravierendes Problem unserer Zeit: Niemand schreibt über das, was nicht ist. Dabei wäre es oft wichtig. Auf puenktlichkeit.ch soll der Benutzer auch sehen, wo es keine Muster gibt. Denn keine Nachricht ist auch eine Nachricht. Und häufig ist es sogar «good news».

Inhaltsverzeichnis

1	Einführung	4
2	Ausgangslage: Vorgängerversion von puenktlichkeit.ch	5
3	Geographische Einordnung	7
4	Verdichtung und Auswahl	10
5	Farbwahl und Codierung	11
6	Zielgruppen und Bedürfnisse	13
7	Und wo ist die «Story»? – Eine Reflektion.	19

1 Einführung

Zu den wichtigsten Erfolgsfaktoren des öffentlichen Verkehrs gehört eine hohe Pünktlichkeit. Diverse Anspruchsgruppen haben folglich ein Interesse daran, die im täglichen Betrieb erzielten Pünktlichkeitswerte erfahren, vergleichen und analysieren zu können: Mitarbeiter der Verkehrsunternehmen, Fahrgäste, Behörden, Medienschaffende und Politiker. Umso erstaunlicher ist es, dass in der Schweiz bislang keine umfassende Pünktlichkeitsstatistik publiziert wird: Zwar veröffentlichen viele der ca. 250 Verkehrsunternehmen aggregierte Einzelwerte (z.B. „Jahrespünktlichkeit“), es existiert aber weder eine öffentliche landesweite Zusammenstellung, noch lassen sich die Werte zeitnah (z.B. Pünktlichkeit der letzten Woche) oder im Detail (z.B. Pünktlichkeit einer bestimmten Linie) einsehen.

Im Sinne der „Open Data Strategie“ des Bundesrats hat das Bundesamt für Verkehr vor einiger Zeit den Auftrag erteilt, Daten der Schweizer Verkehrsunternehmen öffentlich bereit zu stellen. Diese Daten sollen es auch branchenfremden Akteuren ermöglichen, neue Systeme, Apps und Statistiken zum öffentlichen Verkehr zu entwickeln. Erfahrungen im Ausland (z.B. in London) haben gezeigt, dass ein solcher Open Data-Ansatz hohe Innovationskraft freisetzen und zugleich die Verkehrsunternehmen von der Entwicklung eigener Informationssysteme entlasten kann.

Seit Dezember 2016 steht die beauftragte Plattform www.opentransportdata.swiss bereit. In einer ersten Ausbaustufe sind dort u.a. die geplanten (= publizierter Fahrplan) und tatsächlichen Ankunfts- und Abfahrtszeiten (= Verkehrslage) diverser Verkehrsunternehmen abrufbar und dürfen frei genutzt werden. Damit ist die Basis für die Erstellung öffentlicher Pünktlichkeitsstatistiken gegeben.

Ich habe diese Gelegenheit genutzt und im Rahmen einer Semesterarbeit im CAS Business Intelligence eine Auswertungsplattform erstellt: Die Applikation puenktlichkeit.ch liest jede Nacht die neu publizierten Betriebsdaten des Vortages von der Open Data Plattform, überführt sie in ein Data Warehouse und bietet über eine Web-Site die Möglichkeit, diverse Auswertungen zu erstellen. www.puenktlichkeit.ch ist seit April 2017 öffentlich aufgeschaltet und kann von jedermann genutzt werden. Datenmodell, Architektur und Funktionsweise sind ausführlich in meiner damaligen Semesterarbeit beschrieben (siehe Beilage).

Naturgemäss lagen die Schwerpunkte dieser Arbeit auf konzeptionellen Fragen der Datenaufbereitung einerseits und der technischen Umsetzung andererseits. Die Visualisierung der Ergebnisse hatte dagegen nur geringen Stellenwert. Entsprechend präsentierte sich die erste Version von puenktlichkeit.ch: Die Darstellungsmodelle sind suboptimal, Visualisierungen nicht immer intuitiv verständlich, die Bedienung kompliziert, Relevantes kann nur schwer von Irrelevantem unterschieden werden. Die Screenshots im folgenden Abschnitt geben einen Eindruck davon.

Positives Feedback aus der öV-Branche motivierte mich, die Applikation weiter auszubauen und zu verbessern – tatsächlich sind aus der damaligen Semesterarbeit inzwischen kleinere Beratungsmandate bei Verkehrsunternehmen erwachsen. Etwa zeitgleich mit dem Start des CAS Datenvisualisierung habe ich erstmals mit kartografischen Darstellungen der öV-Pünktlichkeit experimentiert. Das Vorhaben erwies sich als gleichermassen spannend und aufwendig, so dass ich mich im Laufe des Semesters entschlossen habe, es zum Gegenstand meines Autorenprojekts zu machen – und andere, zunächst verfolgte Projektideen zu verwerfen.

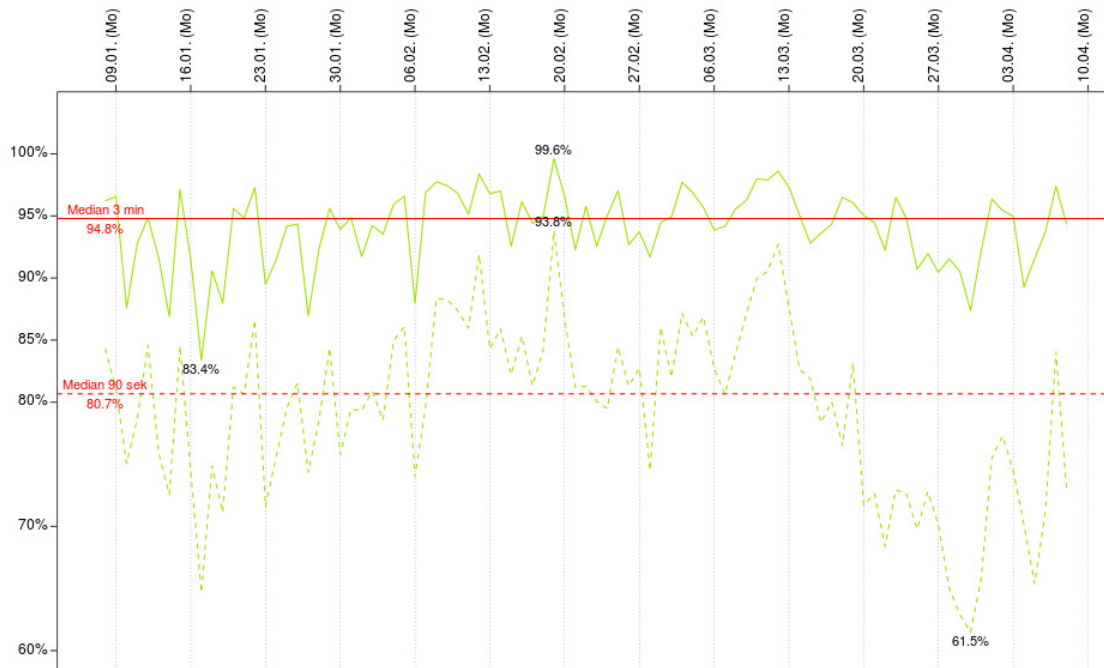
Die Reise von puenktlichkeit.ch V1.0 zu puenktlichkeit.ch V2.0 war eine lange, mit zahlreichen Irrungen und Wirrungen auf allen Ebenen - Daten, Technologie und Visualisierung. Ich konzentriere mich in diesem Bericht darauf, vier ausgewählte Problemfelder näher zu beleuchten (Kapitel 3 bis 6) und schliesse mit einer Reflektion.

Auch die entstandene neue Version von puenktlichkeit.ch ist für Laptops und PC konzipiert: Ideal sind ein grosser Bildschirm und eine Maus mit Zoom-Rad. Grundsätzlich sollte puenktlichkeit.ch auch auf vielen Tablets nutzbar sein – ggf. mit leichten Einschränkungen im Bedienkomfort. Smartphones sind dagegen ungeeignet, da deren Displaygrösse für die von mir gewählte Darstellung nicht ausreicht.

2 Ausgangslage: Vorgängerversion von puenktlichkeit.ch

Die folgenden Screenshots dokumentieren, wie sich Datenvisualisierung und Benutzeroberfläche von puenktlichkeit.ch Anfang April 2017 präsentierten.

Entwicklung von Pünktlichkeitswerten im Zeitverlauf



Parametrisierung der Auswertungen

Indikatoren:

- 5 min: Ankunft <300 sek verspätet
- 3 min: Ankunft <180 sek verspätet
- 90 sek: Ankunft <90 sek verspätet

Betriebstage:

01.03.2017 bis 31.03.2017

< 1T 1W 4W 13W Max >

Verkehrsmittel

Alle

Unternehmen

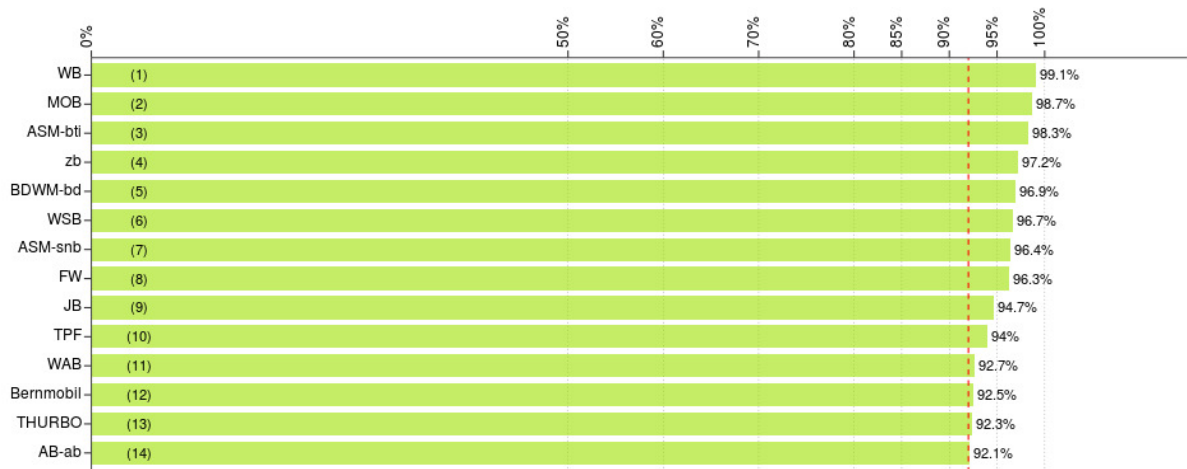
SOB: Schweizerische Südostbahn (SOB)

Kanton SG **Gemeinde** Altstätten

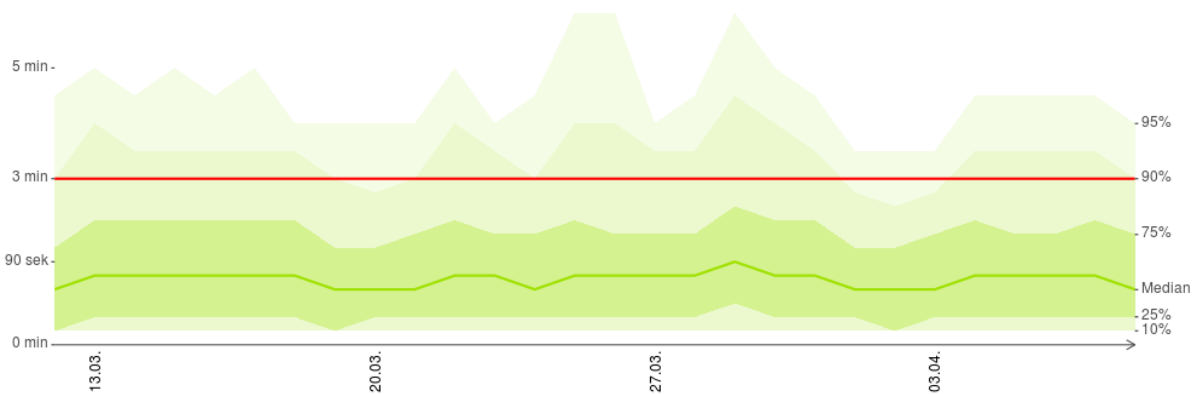
Haltestelle

Alle

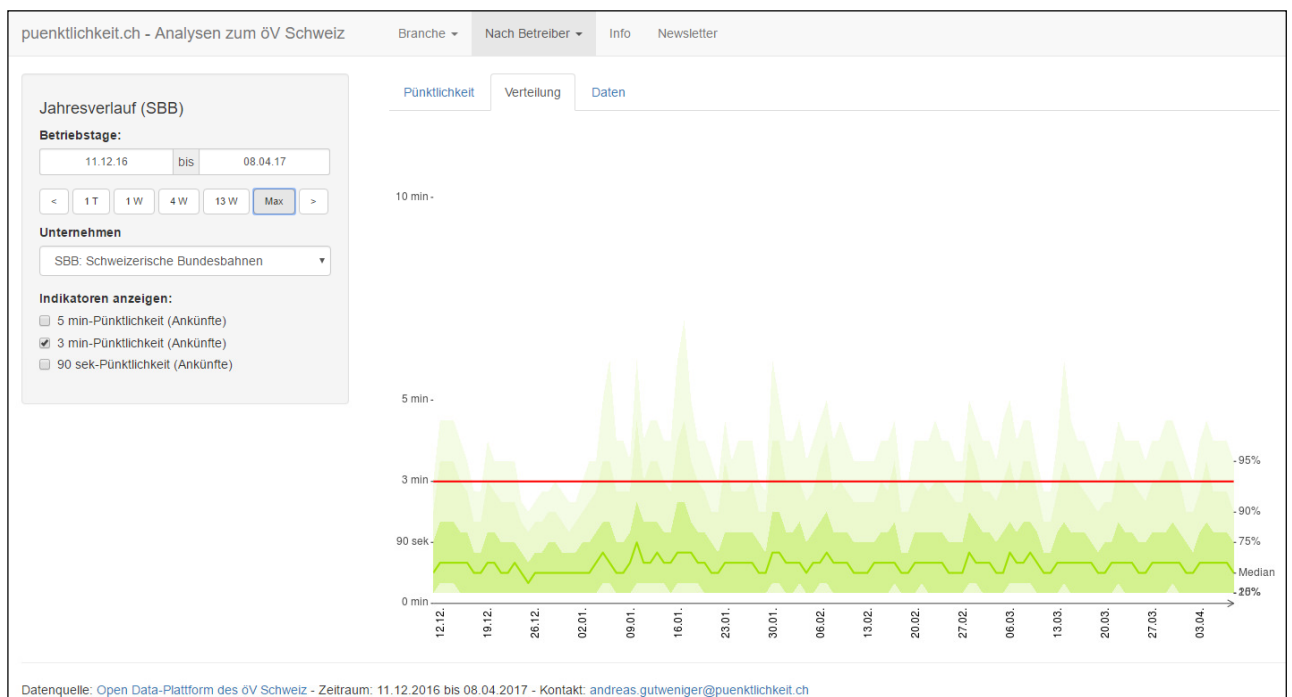
Vergleich der Pünktlichkeitswerte mehrerer Unternehmen



Statistische Verteilung: Welches Verspätungs-Ausmass tritt wie häufig auf?



Informationsarchitektur und Navigationskonzept



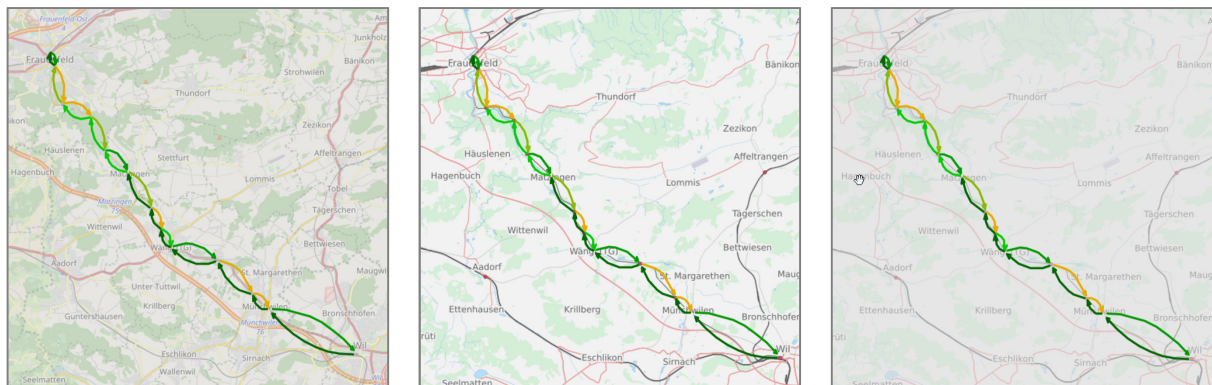
3 Geographische Einordnung

Pünktlichkeitsinformationen haben einen engen Bezug zur Zeit einerseits und zur Geographie andererseits. Ein verbreitetes Visualisierungsmittel für temporale Aspekte ist der (vertikal oder horizontal verlaufende) Zeitstrahl. Davon wurde bereits bei der Vorgängerversion Gebrauch gemacht.

Als anspruchsvoller erweist sich die Geographie: Diese ist in meinem Anwendungsfall zweidimensional (Längen- und Breitengrad; Höhe / Steigungen sind nicht gänzlich irrelevant, aber vernachlässigbar). Und noch wichtiger als die relative Anordnung der Elemente zueinander ist ihre absolute Verortung (der Benutzer braucht Referenzpunkte, z.B. «Bahnhof Bern»). Die Verwendung einer Landkarte ist naheliegend, die Umsetzung aber nicht trivial:

1. Für die Erstellung von eigenem Kartenmaterial fehlen die Ressourcen. Es muss auf Bestehendes (und Verfügbares) zurückgegriffen werden.
2. Gegebenheiten mit Relevanz für den öffentlichen Verkehr sollten gut erkennbar sein, Irrelevantes (z.B. Autobahnen) sollte möglichst wenig Aufmerksamkeit einnehmen.
3. Die Darstellung muss in sehr unterschiedlichen Massstäben funktionieren – von der Totalansicht Schweiz bis zum Fahrtabschnitt zwischen zwei Bushaltestellen.
4. Referenzpunkte müssen auffindbar sein, dürfen aber nicht von der Kernaussage ablenken.

Zahlreiche Karten-Anbieter wurden evaluiert¹. Exemplarisch hier drei der getesteten Konstellationen: Das Beispiel links (OpenStreetmap) enthält zu viel Strassenverkehr. Das Beispiel in der Mitte («Transport» von Thunderforest) fokussiert auf den öV (alle Haltestellen und Linienverläufe sind enthalten), jedoch stören die Farbtöne die farbliche Hervorhebung der Pünktlichkeitsinformation. Gewählt habe ich schliesslich die Variante rechts – die Thunderforest-Karte mit geringerer Deckkraft (Opacity).



Aus diesen Beispielen ist bereits die Lösung zu einem anderen gestalterischen Problem ersichtlich, das mich sehr intensiv beschäftigt hat: Was ist das Bezugsobjekt der Pünktlichkeit? Eine naheliegende Wahl sind Haltestellen, weil sie sich einfach als Punkte auf der Karte verorten lassen – nachstehend ein Beispiel aus einem ähnlichen Projekt anderer Autoren².



¹ Für die eingesetzte Technologie «leaflet» findet sich eine gute Übersicht der verfügbaren Karten-Provider unter <http://leaflet-extras.github.io/leaflet-providers/preview/>

² <http://transportdaten-sbb.opendata.iwi.unibe.ch/> von David Bösiger und Dominik Briner.

Inhaltlich ist das jedoch unbefriedigend, weil in der Haltestellen-Pünktlichkeit die Werte von sehr unterschiedliche Fahrten (zwei Fahrrichtungen pro Linie, an zentralen Orten sehr viele Linien) aggregiert werden. Eine Pünktlichkeit von 95% am HB Zürich ist wenig aussagekräftig. Sowohl Fahrgäste als auch Transportunternehmen möchten es genauer wissen. Es sollte daher die Pünktlichkeit pro Fahrabschnitt dargestellt werden – auf jeden Fall richtungsgetrennt und möglichst auch differenziert nach Verkehrsmittel und Unternehmen. Pfeile scheinen das visuelle Mittel der Wahl, sind aber nicht ohne Schwierigkeiten:

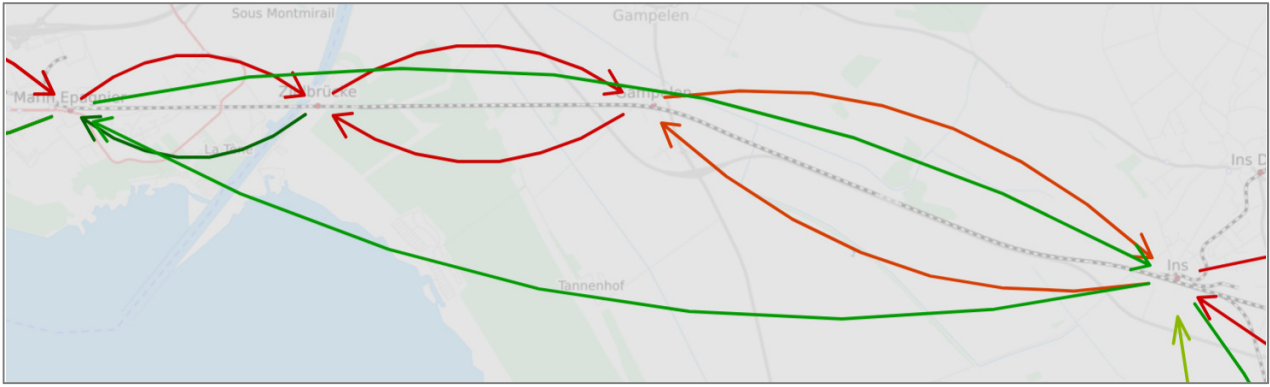
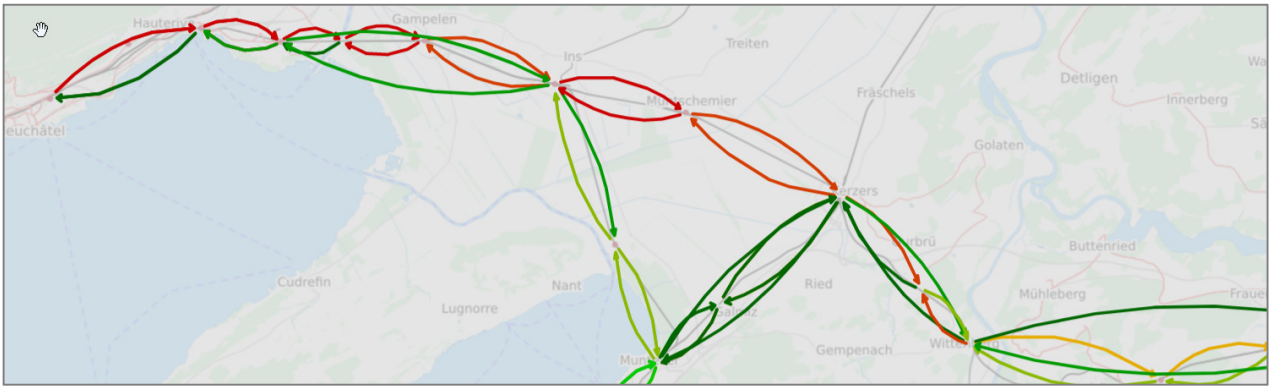
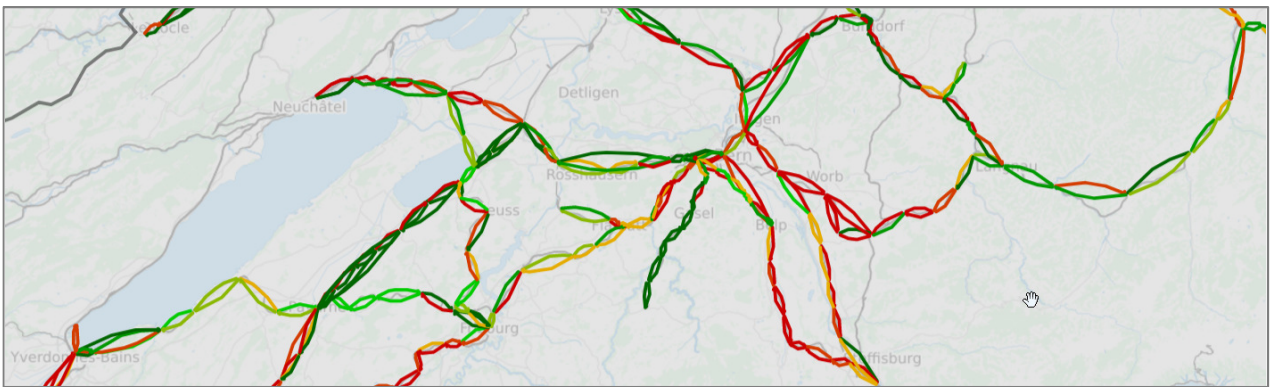
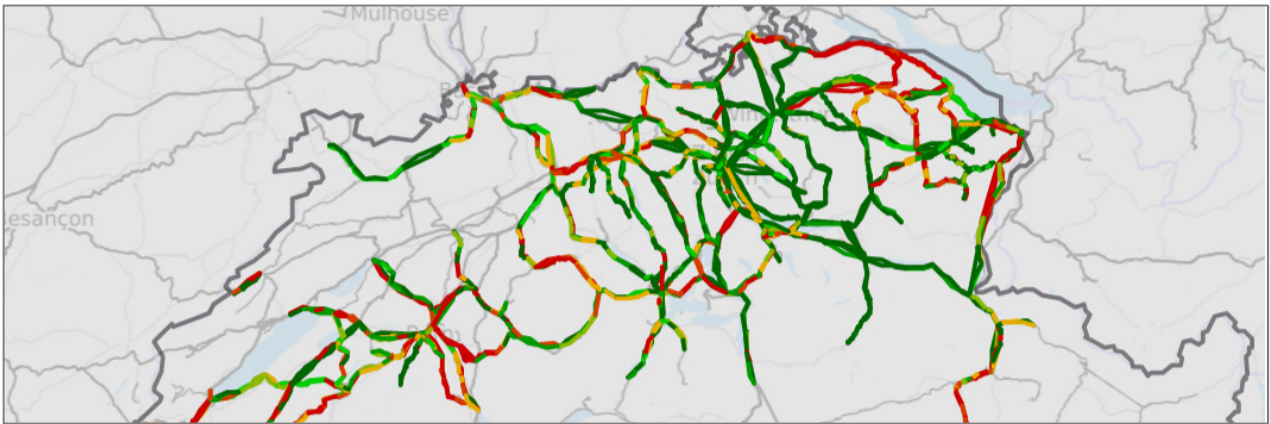
1. Pro Abschnitt werden zwei Pfeile benötigt. Diese müssen geeignet versetzt werden, damit sie nicht überlappen. Dieser Versatz muss in unterschiedlichen Masstäben / Zoomstufen funktionieren, d.h. er darf sich nicht auf fixe Geokoordinaten beziehen.
2. Die Richtung der Pfeile muss erkennbar sein, die Pfeilspitzen sollen aber relevante Karteninformationen möglichst wenig verdecken.
3. Pfeile unterschiedlicher Verkehre müssen klar unterscheidbar sein und sollen sich nicht stören.
4. Zahlreiche Experimente mit geraden Pfeilen führten zu ästhetisch unbefriedigenden Ergebnissen. Neuer Lösungsansatz: Gebogene Pfeile!
5. Das eingesetzte (und von mir sonst überaus geschätzte) Framework «leaflet» bietet kein Darstellungselement für gebogene Pfeile.

Mein Lösungsansatz besteht darin, Pfeile als Polygonzug zu konstruieren, der einer Bezier-Kurve zwischen den beiden zu verbindenden Haltestellen folgt. Dabei werden das erste und das letzte Segment ausgelassen, damit die Anfangs- und Endpunkt nicht verdeckt werden und sich benachbarte Pfeile nicht berühren. Den Abschluss des Polygonzugs bilden zwei Geraden, welche die Pfeilspitze formen.

Das Ergebnis ist nicht völlig befriedigend, stellt aber die beste Lösung der oben beschriebenen Problemsituation dar, die ich finden konnte. Intensiv experimentiert habe ich mit verschiedenen Bogenradien, Pfeillängen und -spitzen. Es stellte sich heraus, dass die Parameter abhängig vom Verkehrstyp gewählt werden sollten, um den sehr unterschiedlichen Haltestellen-Distanzen Rechnung zu tragen (viele Kilometer bei Intercity-Zügen, wenige 100m bei Bussen und Trams).

Die folgende Seite zeigt vier verschiedene Zoomstufen. Dargestellt ist jeweils ein Ausschnitt aus dem Schweizer S-Bahn-Netz. Drei Dinge fallen auf:

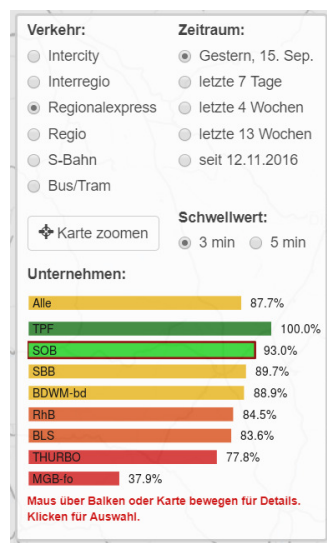
1. Bei kleinem Masstab (vgl. die oberen beiden Abbildungen) sind die Pfeilspitzen nicht mehr zu erkennen. Dies ist einerseits erwünscht, da sie in einer ohnehin recht «dichten» Darstellung eher verwirren als nützen würden. Andererseits ist dadurch die Orientierung nicht mehr zu erkennen. Um dem Benutzer dennoch zu vermitteln, welcher Bogen welche Fahrrichtung repräsentiert, wird eine «intuitive» Anordnung verwendet: Busse und Trams fahren rechts, Züge in der Schweiz normalerweise links. Im S-Bahn-Beispiel repräsentiert also der obere von zwei horizontal angeordneten Bögen eine Fahrt von Westen nach Osten, der untere die Gegenrichtung. Beim Verkehrstyp «Busse und Trams» wäre es umgekehrt.
2. Bei grossem Masstab (vgl. die unterste Abbildung) erscheinen die Bögen «eckig»: Man erkennt, dass es sich um Polygonzüge und nicht um echte Bögen handelt. Dies könnte vermieden werden, indem die Polygonzüge feingranularer berechnet werden. Allerdings erhöht dies die zu übertragende Datenmenge und geht zu Lasten der Ladegeschwindigkeit (allein Bernmobil umfasst 650 Bögen, von denen jeder derzeit aus 10 Segmenten besteht). Die gewählte Granularität stellt somit einen Kompromiss zwischen Ästhetik und Bedienkomfort dar.
3. Für nahe beieinanderliegende Bögen würde man sich fallweise eine andere Anordnung wünschen, um Überlappungen zu vermeiden. In der untersten Abbildung z.B. sollte der obere grüne Bogen weiter nördlich verlaufen und Abstand zu den drei roten Bögen halten. Dies ist der etwas simplen automatischen Anordnung der Elemente geschuldet – ein fortgeschrittener Placement-Algorithmus wäre wünschenswert gewesen, jedoch fehlten die Ressourcen. Immerhin: Im Bild darüber funktioniert die Anordnung der 6 dunkelgrünen Bögen zwischen Muntelier und Kerzers recht gut.



4 Verdichtung und Auswahl

Der Umfang des Quelldatensatzes ist gewaltig: Täglich 260'000 neue Einträge – mittlerweile sind es 75 Millionen. Dies wirft immer wieder die Frage nach dem angemessenen Grad der Verdichtung auf: Aggregiere ich zu stark, gehen interessante Details verloren; aggregiere ich zu wenig, ist die Darstellung nicht mehr interpretierbar.

Auch hierzu habe ich diverse Versuche angestellt – und mich schliesslich dafür entschieden, alle jene Fahrten zu einem Pfeilbogen zusammenzufassen, die zwischen denselben Haltestellen liegen, zum gleichen Unternehmen gehören und der gleichen Verkehrsart zuzuordnen sind. Das heisst konkret: Je ein Bogen für Intercity (SBB), Interregio (SBB) und Regio (RA) von Brig nach Visp. Aber ein gemeinsamer Bogen für die Berner Buslinien 11 und 21 vom Henkerbrännli zum Bierhübeli.



Ergänzend zur Visualisierung kann Interaktivität eingesetzt werden, um den Benutzer bei der Komplexitätsbewältigung zu unterstützen. «Unterstützen», das heisst: ihm sinnvolle Auswahlmöglichkeiten eröffnen; nicht aber, ihn mit «anything goes» zu überfordern.

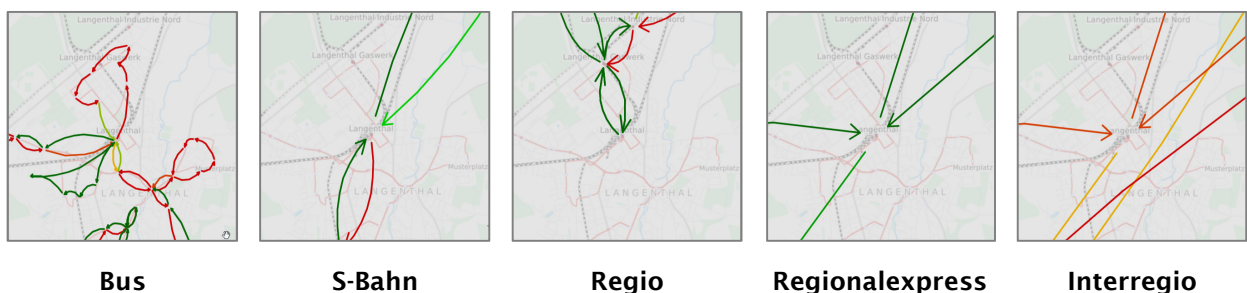
Gegenüber der Vorgängerversion habe ich die Auswahlmöglichkeiten daher eingeschränkt. In der Karten-Darstellung kann der Benutzer Folgendes einstellen:

- Art des Verkehrs
- Zeitraum
- Schwellwert für Pünktlichkeit

Ergänzend hat er die Möglichkeit, die Darstellung auf ein einzelnes Unternehmen einzuschränken (durch Klick auf den entsprechenden Balken).

Und schliesslich kann er mit Maus und Zoom-Rad den Ausschnitt und den Massstab der Karte wählen.

Mit 6 Ausprägungen bewegt sich die Verkehrsart am oberen Rand des Sinnvollen: weniger wäre durchaus wünschenswert, zumal längst nicht alle Benutzer die verwendeten Zugskategorien gut unterscheiden können. Eine derart breite Differenzierung war aber notwendig, um eine Überfrachtung der Landkarte mit vielen unterschiedlichen Verkehren zu vermeiden, wie das Beispiel Langenthal zeigt:



Vor allem aber soll mit dieser Unterscheidung sichergestellt werden, dass auf puenktlichkeit.ch nur Vergleichbares miteinander verglichen wird: Die Erwartungen, aber auch die betrieblichen Herausforderungen unterschieden sich stark zwischen den Verkehrsarten – und folglich auch die erzielten Pünktlichkeitswerte. Es ist daher nicht seriös, eine S-Bahn und einen Interregio vergleichen zu wollen.

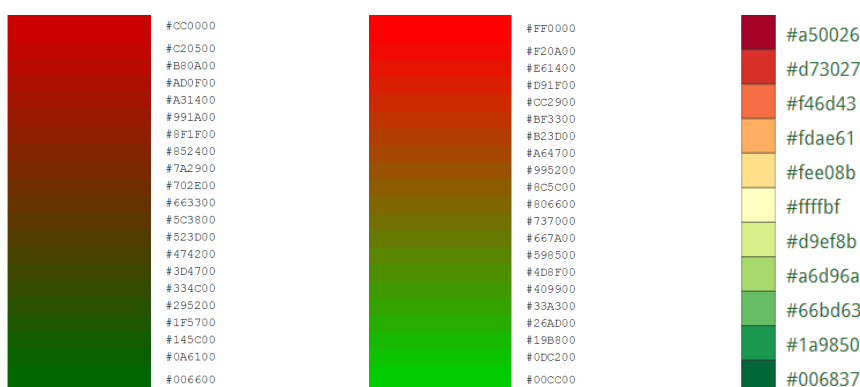
Wie bei jedem Datensatz, so ist es auch hier sinnvoll, Irrelevantes vorab auszufiltern. Im öffentlichen Verkehr gibt es eine Vielzahl von Sonderfällen, die jeder für sich genommen selten auftreten, in Summe die Komplexität aber enorm erhöhen. Ich versuche, dem Benutzer diese Komplexität zu ersparen, indem nur Abschnitte dargestellt werden, die mindestens 5x wöchentlich befahren werden.

5 Farbwahl und Codierung

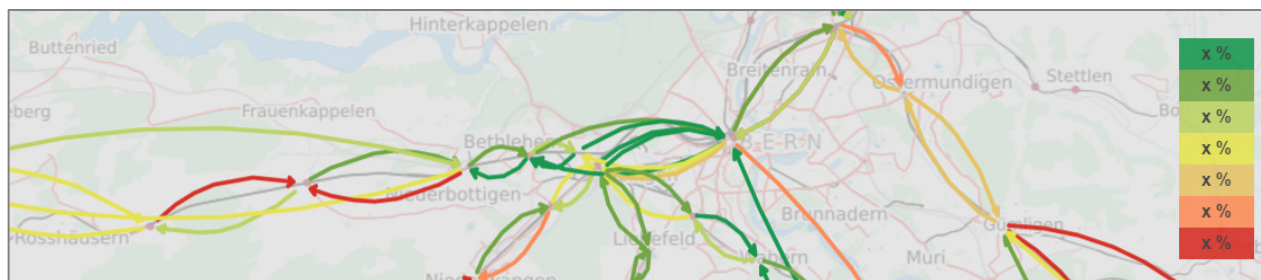
Als weitere grosse Herausforderung erwies sich die farbliche Codierung der Pünktlichkeitswerte. Naheliegender war, hohe Pünktlichkeit grün, niedrige Pünktlichkeit rot darzustellen. Folgende Probleme mussten gelöst werden:

1. Wie ist der Farbverlauf? Welche Zwischentöne werden verwendet?
2. Wie ist die Codierung? Welche Pünktlichkeitswerte werden welchem Farbton der Skala zugeordnet?

Bei Experimenten zum Farbverlauf schied bi-chromatische Skalen (linke und mittlere Abbildung³) recht bald aus: Zwischen rot und grün liegen Schlammtöne, die einerseits unschön anzusehen sind und andererseits schlecht mit dem Karten-Hintergrund kontrastieren. Wenn man dem Ampel-Schema folgt, liegt dagegen gelb in der Mitte – siehe den Vorschlag des «Color-Brewers»⁴ im Bild rechts.



Auch hier stellte sich der Kontrast zum Hintergrund als zu gering heraus, so dass ich es mit kräftigeren Gelb- und Orange-Tönen versuchte.



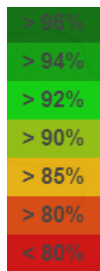
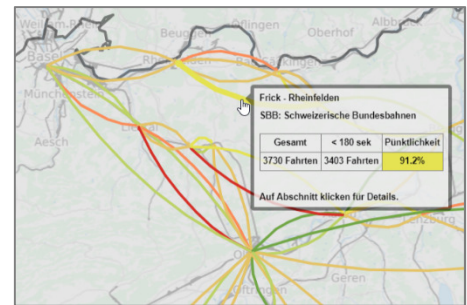
Der Kontrast war nun gut und die Farben differenzierten stark. Zu stark, wie ich fand, denn das Bild wirkte nun sehr bunt. Vor allem aber: Es wurde nicht mehr klar, welche Ordnung die Farben haben und welches die «interessanten» Abschnitte sind: Sticht gelb heraus? Oder orange? Oder rot? Oder doch das kräftige Grün?

Ich habe die Farbtöne mehrmals angepasst, bis ich beim heutigen Zustand ankam. Wichtiger noch: Die Schwellwerte und damit die Auftretenswahrscheinlichkeit der einzelnen Farben mussten justiert werden. Aus statistischer Sicht ist es naheliegend, die Codierung so zu wählen, dass alle Farben gleich häufig auftreten. Dies erreicht man durch Wahl der entsprechenden Perzentilwerte aus der Häufigkeitsverteilung - so bin ich auch im oben abgebildeten Beispiel vorgegangen. Zur Vermittlung einer pointierten Aussage erweist sich das als ungeeignet: die Zwischentöne lenken ab und sollten daher nur selten auftreten.

³ Quelle. https://www.w3schools.com/colors/colors_mixer.asp

⁴ Quelle. <http://colorbrewer2.org/#type=diverging&scheme=RdYlGn&n=11>

Viel gravierender: Eine gleichmässige Verteilung der Farben nimmt die Interpretation der Werte vorweg. So ergibt sich für die Fahrt von Frick nach Rheinfelden eine Einstufung als «gelb». Fakt ist: 9 von 10 Zügen sind dort weniger als 180 Sekunden zu spät. Sollte man das wirklich als «gelb» taxieren? Relativ zu anderen Verkehren in der Schweiz ist es offenbar mittelmässig – aber ist das ein angemessenes Kriterium? In den meisten Nachbarländern sind Verspätungen von 3 Minuten nicht einmal eine Erwähnung wert.



Ich habe mich schliesslich für folgende Pünktlichkeits-Skala entschieden:

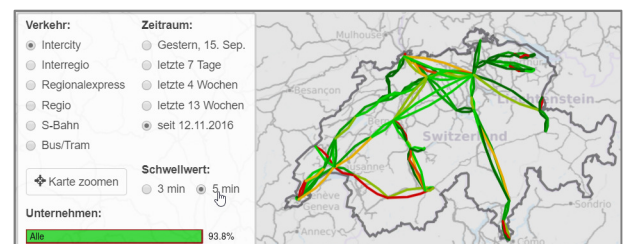
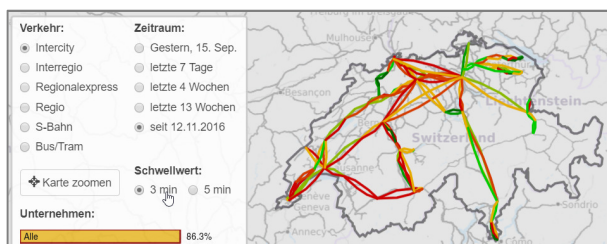
- oberhalb 90%: grün (abgestuft in vier Farbtöne)
- 85-90%: orange (gelb kommt aus oben genannten Gründen nicht mehr vor)
- Unterhalb 85%: rot (abgestuft in zwei Farbtöne)

Dies ist natürlich eine subjektive Festlegung. Im internationalen Vergleich scheint sie mir nach wie vor streng. Zugleich entspricht sie meiner Wahrnehmung, dass der öV in der Schweiz grundsätzlich recht «grün» unterwegs ist – mit einigen Farbtupfern.

Wie ich in Kapitel 4 beschrieben habe, kann der Benutzer zwischen zwei «Pünktlichkeits-Niveaus» auswählen: Liegt der Schwellwert für eine unpünktliche Ankunft bei 3 oder bei 5 Minuten?

Natürlich beeinflusst diese Wahl die berechneten Pünktlichkeitswerte stark – bei 5 Minuten werden viel weniger Fahrten als verspätet taxiert. «Fair» wäre es daher, zwei unterschiedliche Skalen für die beiden Niveaus zu verwenden. Ich habe dies versucht – und dann wieder verworfen. Denn die Effekte sind für statisch ungeübte Benutzer wenig nachvollziehbar: Warum kann ein Verkehr, der beim 3 Minuten-Niveau «grün» ist, im 5 Minuten-Niveau «orange» sein?

puenktlichkeit.ch verwendet nun einheitlich die oben angegebene Skala. Bei allen Pünktlichkeits-Niveaus, bei allen Verkehrsarten, in allen Auswertungen. Mit einer sympathischen Konsequenz: Wer sich und dem öV 2 Minuten mehr Zeit gibt, für den wird die Aussicht deutlich grüner.



6 Zielgruppen und Bedürfnisse

puenktlichkeit.ch richtet sich an ein breites Publikum: vom Pendler, der wissen will, ob sich seine Pünktlichkeits-Wahrnehmung mit der Statistik deckt, über den fachlich interessierten Laien und den ferrophilen Enthusiasten (aka «Puffer-Knutscher») bis hin zu den Experten der Verkehrsunternehmen. Nicht zu vergessen: Politik, Verwaltung und Medien.

Diese Gruppen haben unterschiedliche Informations-Bedürfnisse, unterschiedliche fachliche Vorkenntnisse und vor allem: unterschiedlich viel Erfahrung im Umgang mit statistischen Auswertungen. Dies gilt es zu berücksichtigen.

Zu den Zielgruppen stellt sich auch noch eine andere Frage: Liest dies hier überhaupt irgendjemand? Um es zu erfahren, schenke ich den ersten drei Personen, die sich bei mir noch im 2017 mit der Lösung «pünktlich gelesen» melden, einen öV-Gutschein. E-Mails an andreas.gutweniger@outlook.com.

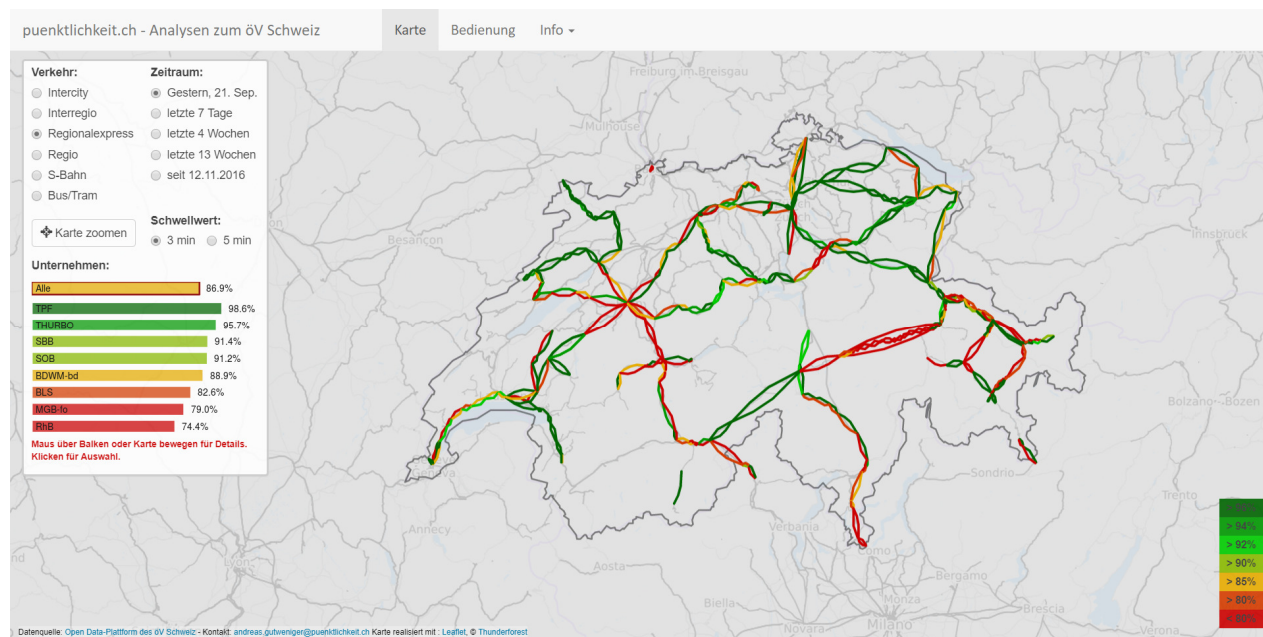
Trotz allen Unterschieden darf eine Gemeinsamkeit vermutet werden: Die schrittweise Verfeinerung vom Groben ins Detail ist angemessen, um die gewünschte Information aufzufinden. puenktlichkeit.ch ist so aufgebaut, dass der Benutzer auf der Reise vom «Big Picture» zum einzelnen Datenpunkt unterstützt wird – ohne ihm diese Abfolge explizit vorzugeben. Parallel zum Detailierungsgrad nimmt dabei auch der Anspruch der Auswertungen zu: Der Benutzer kann dort aufhören, wo es ihm zu viel wird – zu viel Detail oder zu viel Schwierigkeit. Und einen Gewinn haben bis zum Erreichen dieses Punkts.

Auch wenn sie nicht streng sequentiell durchlaufen werden müssen, lassen sich grundsätzlich folgende Phasen unterscheiden:

1. Überblick
2. Einschränkung
3. Navigation
4. Auswahl
5. Perspektive
6. Datenpunkt
7. Interpretation

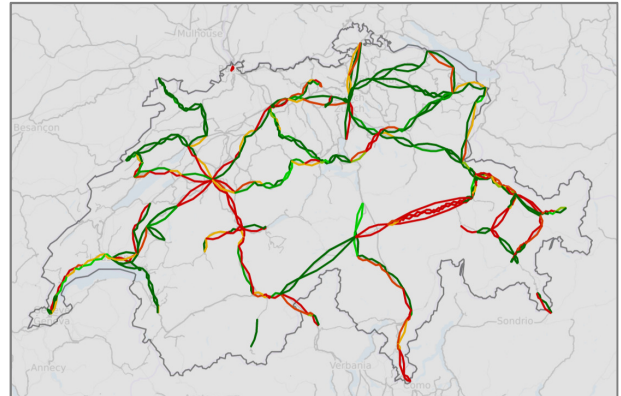
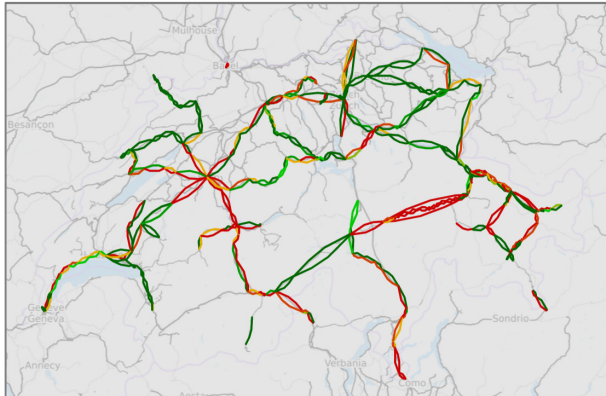
Phase 1: Überblick

Der Einstieg in die Applikation erfolgt über die Gesamtansicht der Schweiz. Links davon alle Bedienelemente, rechts die Farblegende, oben eine kompakte Menüleiste, unten Hinweise zur Urheberschaft.



Als Zeitraum ist der Vortag eingestellt – damit dies besser bemerkt wird, ist er mit dem jeweiligen Datum angeschrieben. Diese Grundeinstellung führt dazu, dass die Karte jeden Tag anders aussieht. Und doch dürfte der erste Eindruck an den meisten Tagen ähnlich sein: Grün überwiegt - der öV Schweiz ist grundsätzlich pünktlich. Zugleich laden die unterschiedlichen Farben dazu ein, es etwas differenzierter anzuschauen.

Bis zum wertvollen Hinweis eines Kurs-Kollegen war meine Schweiz konturlos, das Anwendungsgebiet der Applikation allein durch die farbigen Verkehrslinien zu erkennen.



Der vorstehende Vergleich vermittelt eine Erkenntnis, die mich schmerzt, auch wenn sie nicht neu ist: Grenzen sind ein Ur-Bedürfnis - indem sie ab-grenzen, verschaffen sie Orientierung. Hier zeigen sie vor allem, worum es geht: den öffentlichen Verkehr in der SCHWEIZ. puenktlichkeit punkt CH.

Phase 2: Einschränkung

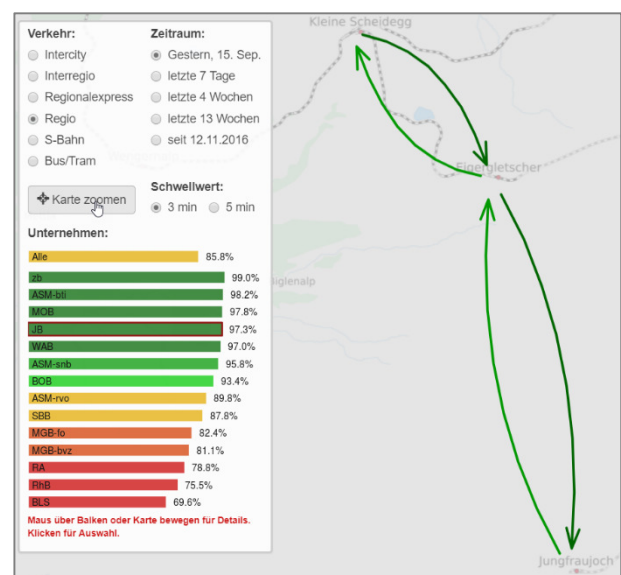
Der Benutzer kann sein Interesse konkretisieren: Welche Verkehrsart? Welcher Zeithorizont? Welches Unternehmen interessiert ihn? Welchen Massstab legt er bei der Pünktlichkeit an – maximal 3 oder maximal 5 min später als der Fahrplan?

Die wichtigsten Design-Entscheidungen zu dieser Phase habe ich bereits in Kapitel 4 beschrieben.

Phase 3: Navigation

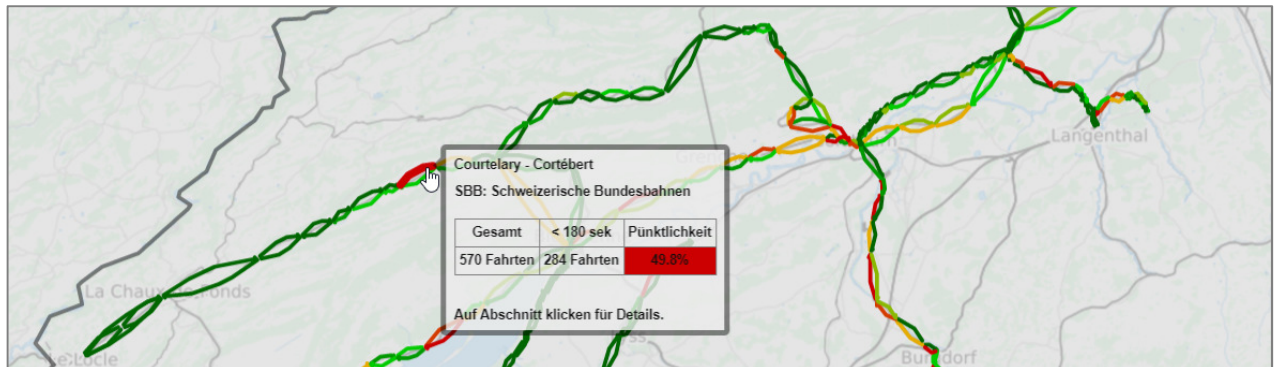
Auch hier trifft der Benutzer eine Wahl: Welcher Teil der Schweiz ist von Interesse? Interessiert eine Region, ein Ort, eine bestimmte Verkehrslinie? Zur Verfügung stehen die von anderen Web-Karten gewohnten Navigationsmöglichkeiten zum Verschieben und Zoomen.

Und weil die in Phase 2 getroffene Wahl möglicherweise bereits einen geographischen Ausschnitt repräsentiert, gibt es einen Button, mit dem der Kartenausschnitt automatisch an die ausgewählten Verkehre und Unternehmen angepasst wird: Wen die Pünktlichkeit der Jungfraubahn interessiert, der kann mit einem Klick dorthin gelangen. Und ebenso einfach zurück, wenn sich das Interesse geändert hat.



Phase 4: Auswahl

Die Kartendarstellung gibt dem Benutzer einen Überblick: Welcher Abschnitt interessiert mich? Ein Mouseover liefert erste Zusatzinformationen.



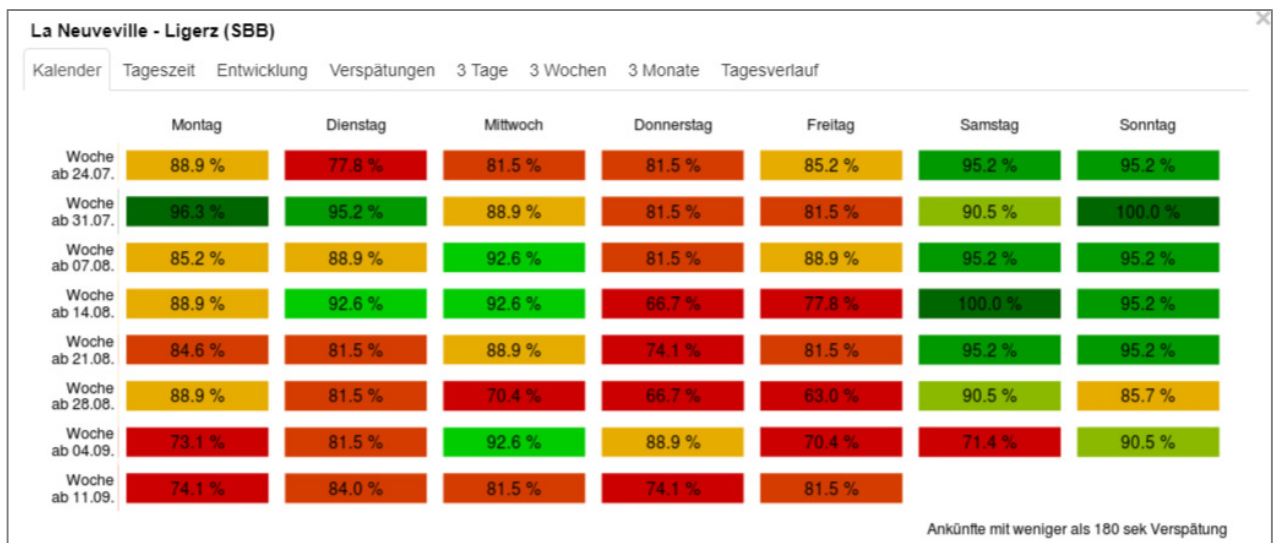
Mit einem Klick auf den Abschnitt taucht er ab in die Details.

Phase 5: Perspektive

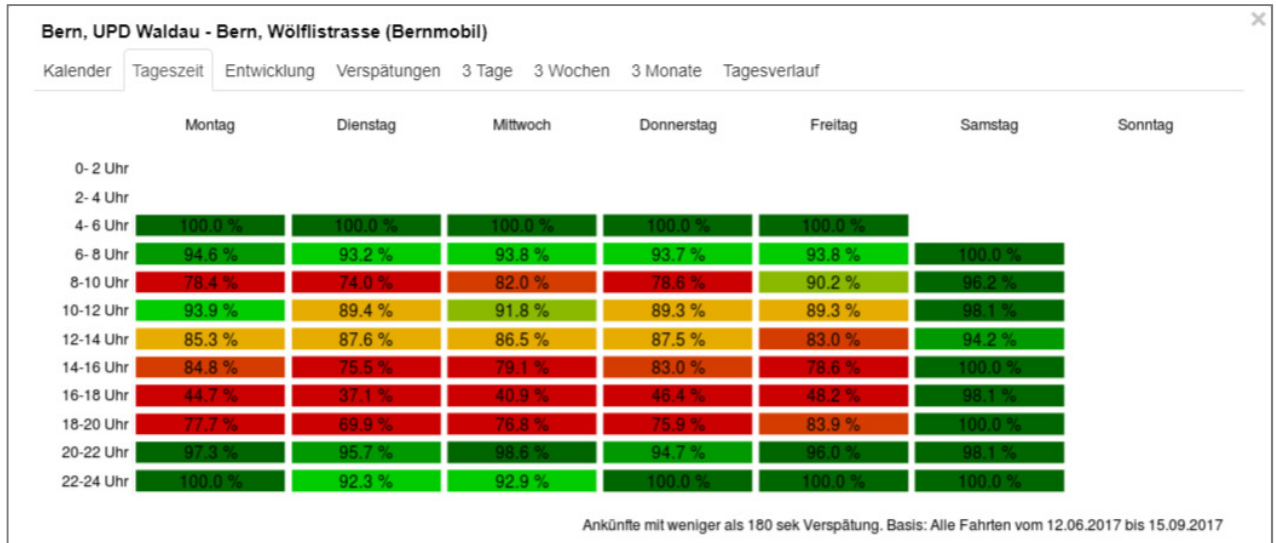
Spätestens bei der Detailanalyse laufen die Bedürfnisse der Benutzer stark auseinander. Und weil es keine Darstellung gibt, die eine Antwort auf alle relevanten Fragen liefert, werden hier verschiedene Perspektiven zur Auswahl angeboten.

Zu jeder Perspektive gibt es eine Registerkarte (Tab) mit passender Visualisierung. Die Anordnung richtet sich nach der bekannten Reihenfolge - vom Groben ins Detail, vom Laien-Niveau zum Experten: Links sind die intuitiv verständlichen Darstellungen, je weiter man nach rechts kommt, desto mehr Fachkenntnis, Interpretation und statistisches Verständnis wird benötigt, desto mehr Details werden aber auch dargestellt.

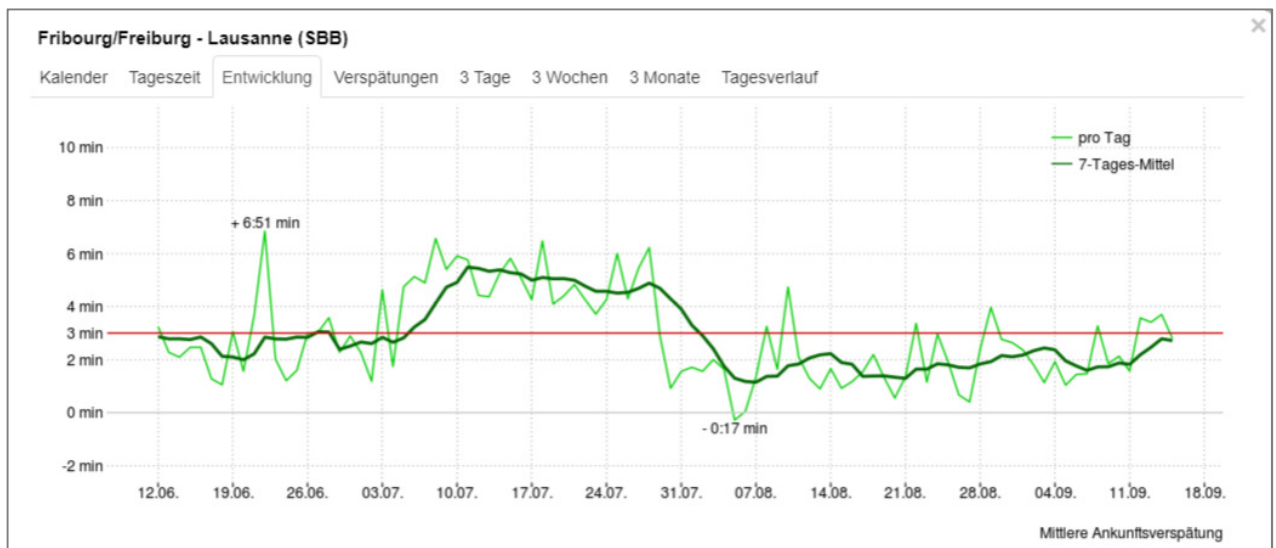
An welchen Tagen gab es Verspätungen?



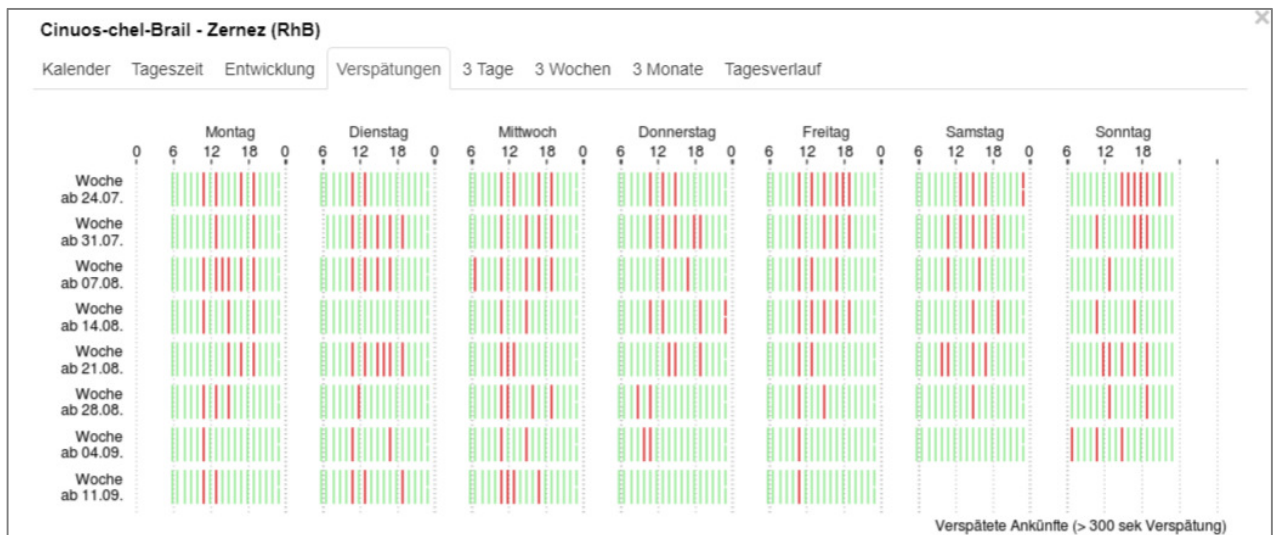
Und zu welchen Tageszeiten?



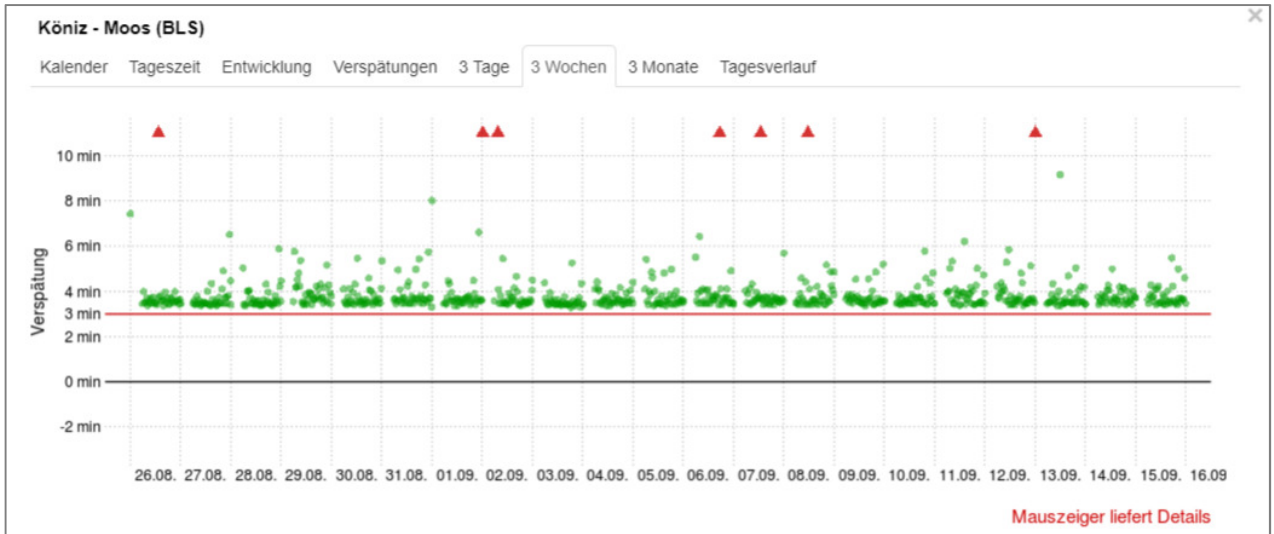
Gibt es einen Trend?



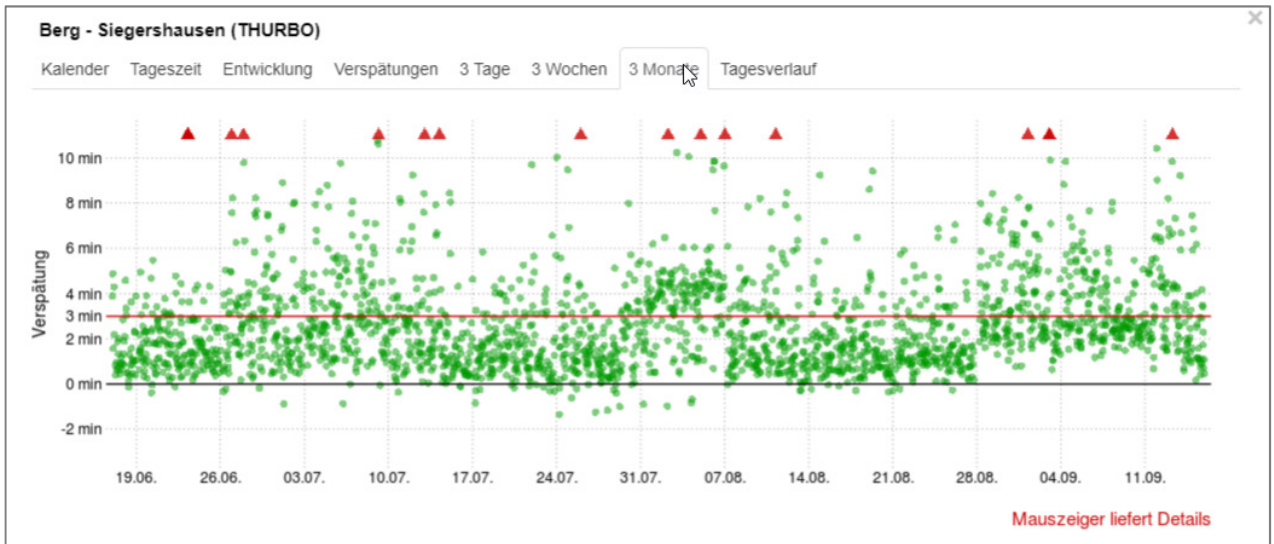
Oder gar ein Muster?



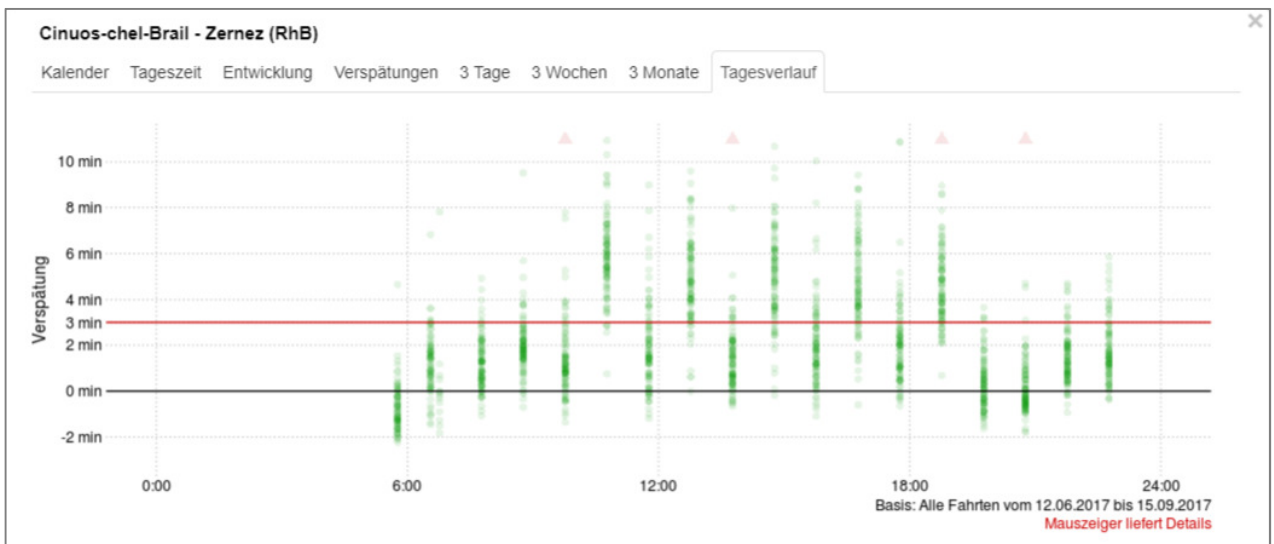
Welches Ausmass hatten die Verspätungen in den letzten 3 Tagen, 3 Wochen, ...



... 3 Monaten?

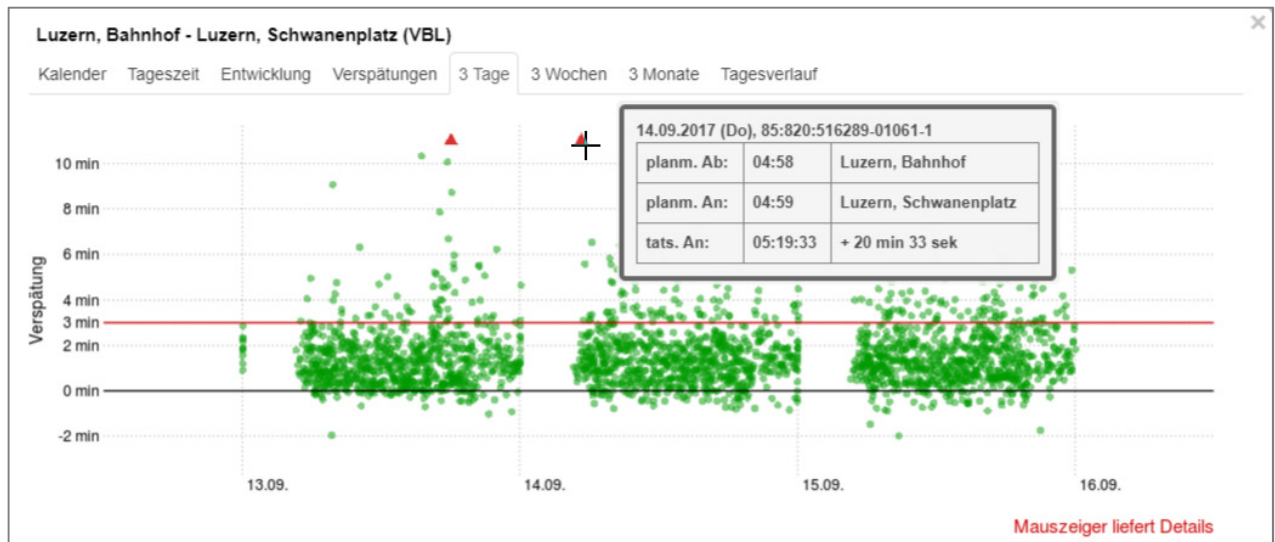


Und im Tagesverlauf?



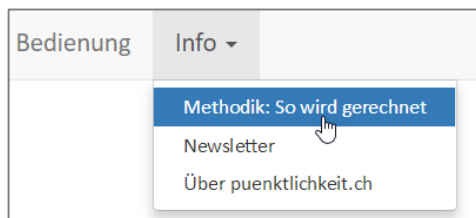
Phase 6: Datenpunkt

Die letzten vier Perspektiven (3 Tage, 3 Wochen, 3 Monate, Tagesverlauf) visualisieren jeden einzelnen Halt des betrachteten Abschnitts. Der Benutzer ist auf der Ebene «Datenpunkt» angekommen. Aber welcher ist welcher? Speziell bei Extremwerten besteht oft das Bedürfnis, sie identifizieren zu können. Adressiert wird es durch eine Mouseover-Funktion.



Phase 7: Interpretation

Die Interpretation von Daten ist nicht möglich ohne ein fundiertes Verständnis von Quelle und Kontext. Auf puenktlichkeit.ch gibt es daher eine ausführliche Beschreibung der verwendeten Methodik.



Es wird erklärt

- welche Daten verwendet werden,
- aus welcher Quelle sie stammen,
- wie daraus Pünktlichkeitswerte berechnet werden,
- was beim Vergleich der Werte zu beachten ist,
- warum andere Statistiken zu anderen Werten kommen.

7 Und wo ist die «Story»? – Eine Reflektion.

Ich habe gelitten in diesem CAS. Spätestens seit dem Modul von Julian Schmidli zum Datenjournalismus – das meiner Meinung nach eines der besten war.

Storytelling. Geschichten erzählen mit Daten. Die Message rüber bringen. Wer will das nicht?

In einer Welt der Reiz- und Informationsüberflutung sehnen wir uns nach klaren, prägnanten Aussagen. Informationsgrafik soll es auf den Punkt bringen, das Hintergrund-Rauschen des Irrelevanten entfernen. Die Adressaten sollen nicht gelangweilt werden mit jenen Fakten, die keine «Story» haben. Denn das würde ihre Zeit vergeuden – und die Reputation des Absenders schädigen.

Das ist ein legitimer Anspruch – nicht nur für Journalisten, Werber und PR-Leute. Auch als Informatiker muss ich auf den Punkt bringen können, was ich geschaffen, herausgefunden, verbessert habe. Und warum das wertvoll ist für mein Unternehmen.

Ich habe in diesem CAS viel darüber gelernt, wie man eine Geschichte visuell auf den Punkt bringt. Wie man mit Infografiken Fakten beschreiben, interpretieren, aber auch verfälschen kann. Das ist überaus nützlich. Doch was passiert, wenn die Prägnanz der Geschichte zum Mass aller Dinge wird?

Es ist noch nicht lange her, da konnte beeindruckend, wer in seinem Vortrag ein Powerpoint mit Balken- oder Tortendiagrammen zeigte. Seither haben sich die Möglichkeiten der Datenvisualisierung rasant entwickelt. Und jede neue Verbesserung findet rasche Verbreitung. Erkenntnis durch Visualisierung. Ein unschätzbare wertvoller Motor des Fortschritts.

Und zugleich reichlich Kriegsmaterial für das visuelle Wettrüsten. Denn in der Aufmerksamkeitsökonomie sind der Inflation der Geschichten keine Grenzen gesetzt. Da wird statistisches Rauschen nicht beim Namen genannt, sondern visuell aufgemotzt. Der Autor braucht eine Story.

Mein Datensatz umfasst 85 Millionen Einträge. Genug Material, um fast jede These stützen zu können. Bergauf sind Züge pünktlicher als bergab. Die Einschaltquote von Netflix korreliert mit den Verspätungen am Bahnhof X. Rote Busse sind pünktlicher als blaue. Wer die Daten gut auswählt und geschickt visualisiert, kann das belegen. Und sonst halt das Gegenteil. Der Autor braucht eine Story.

Aber Fortschritte in der Visualisierung ändern nichts an der Entropie des Universums. «Patterns» sind selten in den Daten. Rauschen überwiegt. Das ist auch in meinem Datensatz so. Denn sind wir ehrlich: Verspätungen passieren aus den unterschiedlichsten Gründen. Hier eine defekte Weiche, dort ein Personenunfall, hier eine Schulreise, dort ein Blitzschlag. Das meiste davon völlig unsystematisch.

Für mein Projekt habe ich einen generischen Ansatz gewählt: Die Auswertungen sind parametrisierbar, jeden Tag kommen neue Daten. Ob das Ergebnis eine Story liefert, kann man nicht vorab wissen. In den vorausgegangenen Kapiteln habe ich mich um Beispiele bemüht, die eine Aussage transportieren. Es ist mir wichtig zu betonen, dass dies Ausnahmen sind – der Regelfall ist viel langweiliger.

Ein Beispiel: Der Intercity von Zürich nach Bern gehört mit einer 3-Minuten-Pünktlichkeit von nur 80.7% zu den Problemkindern der SBB. Und sehr viele Menschen würden das gern ändern. Nutzen wir also puenktlichkeit.ch, um nach «Patterns» zu suchen:



Die ernüchternde Erkenntnis ist: Hier sind keine Patterns! (auch nicht bei grösserer Auflösung)
Sorry Julian, keine Story!

Wirklich?

«Publication Bias» ist ein gravierendes Problem unserer Zeit: Niemand schreibt über das, was nicht ist. Dabei wäre es oft wichtig. Millionen von Patienten werden mit nutzlosen Therapien behandelt – weil niemand publiziert hat, wann sie NICHT wirken. Und die Challenger-Katastrophe hätte nicht stattgefunden, wenn man ausser den «interessanten» auch die «langweiligen» Datensätze analysiert hätte⁵.

«Keine Nachricht ist auch eine Nachricht» sagt man. Englisch sogar: «No news is good news». Für die Pünktlichkeit im öV Schweiz gilt Letzteres. Denn was würde es bedeuten, wenn sich bei der wichtigsten Intercity-Verbindung des Landes ein Verspätungs-Muster finden liesse? Man müsste daraus schliessen, dass es eine systematische Störung gibt, die bisher noch nicht behoben wurde. Futter für Enthüllungsjournalisten – und ein Armutszeugnis für die SBB!

Was wir in den Auswertungen stattdessen sehen, ist die Abwesenheit von Mustern. Zwischen Zürich und Bern ist nur Rauschen: reiner Zufall, unvorhersehbar. Das spricht für ein sehr weit optimiertes System. Ist das nicht auch eine Grafik wert?

Züge transportieren Passagiere. Infografiken transportieren Aussagen. Züge fahren auch ohne Passagiere. Und Infografiken?

⁵ <https://tamino.wordpress.com/2011/12/14/the-value-of-data/>