

Studienarbeit Informationsvisualisierung

Corona Visualisierung von Ulrich Stark

Data Transformation

Das Datenset, das als Grundlage für die Werte in der Visualisierung benutzt wird, habe ich von <https://ourworldindata.org/coronavirus-source-data> heruntergeladen. Der Grund für die Wahl dieses Datensets war, dass es eines der ersten Ergebnisse bei der Google-Suche nach „covid 19 data“ war und einen vielversprechenden Umfang besaß. Es liegt als CSV-Datei im Unterordner /data vor.

Zuerst musste ich mir einen Überblick über den Umfang des Datenset verschaffen, indem ich es mit dem Framework Pandas einlese und analysiere. Datenspalten schien es mehr als genug zu enthalten. Darunter auch viele Attribute eines Landes, die meine Visualisierung nicht benötigt. Beispielsweise Attribute wie Anteil an Raucher pro Geschlecht oder Lebenserwartung der Bevölkerung, die bestimmt für manche Visualisierungen Sinn machen. In meinem Fall brauche ich nur die Kennung und den vollen Namen eines Landes und die Zahl der Neuinfizierten pro Tag.

Im nächsten Schritt wollte ich sichergehen, dass alle Staaten der Welt in diesem Datenset enthalten sind. Eine Ausgabe der verschiedenen Werte der „location“ Spalte lieferte 212 Regionen zurück. Zurzeit existieren um die 195 unabhängige Staaten auf der Welt. Damit sollten alle Staaten in dem von mir ausgewählten Datenset enthalten sein. Die Differenz der beiden Zahlen ist wahrscheinlich darauf zurückzuführen, dass einige Regionen im Datenset nicht als unabhängiger Staat angesehen werden.

Daraufhin habe ich mit dem Framework matplotlib den zeitlichen Verlauf der Neuinfektionen von Deutschland und USA visualisiert und verglichen. Deutschland scheint sich seit einem klar erkennbaren Anstieg der Neuinfektionen in einer stetigen Abnahme eben dieser Kennzahl zu befinden, im Gegensatz dazu folgt bei den USA nach dem ersten Anstieg ein zweiter noch stärkerer Anstieg, ohne dass die Anzahl der Neuinfektionen dazwischen erkennbar abgenommen hätten. Dieser drastische Unterschied zwischen beiden Kurven führte zu meiner Idee, genau diese Veränderung der Neuinfektionen pro Land in meiner Studienarbeit zu visualisieren. Dabei sollen jeweils zwei Wochen mit ihren Neuinfektionszahlen verglichen werden. Visualisierung von Wochen macht mehr Sinn, als Tage zu vergleichen. Denn damit kann man mögliche Schwankungen der Neuinfektionen zwischen den Tagen ausgleichen und erhält ein aussagekräftigeres und breiteres Bild von der Entwicklung dieser Kennzahl.

Vor dem Exportieren galt es nur noch, die drei benötigten Datenspalten aus dem Datenset zu extrahieren und alle Einträge zu entfernen, die undefinierte Werte enthielten. Wie zum Beispiel Einträge, die als Länderkennung NaN (Not a Number) besaßen. Das führt zu einem übersichtlicheren Datenset und zu einer kleineren Dateigröße, um die Ladezeit der Webseite für den Benutzer zu verringern. Das Ergebnis des Exportierens liegt wieder als CSV-Datei im Unterordner /web/data vor.

Visual Mapping

Die Visualisierung der Daten wird auf einer Weltkarte realisiert, auf der jedes Land eine bestimmte Farbe erhält. Die Weltkarte ist eine SVG-Vektordatei, die von <https://simplemaps.com/resources/svg-world> heruntergeladen wurde und laut Autor ohne Einschränkungen frei verwendbar ist.

Die Farbe eines jeden Landes befindet sich auf einem Spektrum zwischen Rot über Grau zu Grün. Für die Auswahl einer bestimmten Farbe wird der relative Unterschied zwischen den Neuinfektionen

einer vergangenen Woche und der aktuellen Woche berechnet. Eine Zunahme der Neuinfektionen, also ein positiver relativer Unterschied wird dabei in Rot dargestellt, mit 200% und höheren Werten als maximales Rot. Eine Abnahme der Neuinfektionen, also ein negativer relativer Unterschied wird dagegen in Grün dargestellt, mit -200% und kleineren Werten als maximales Grün.

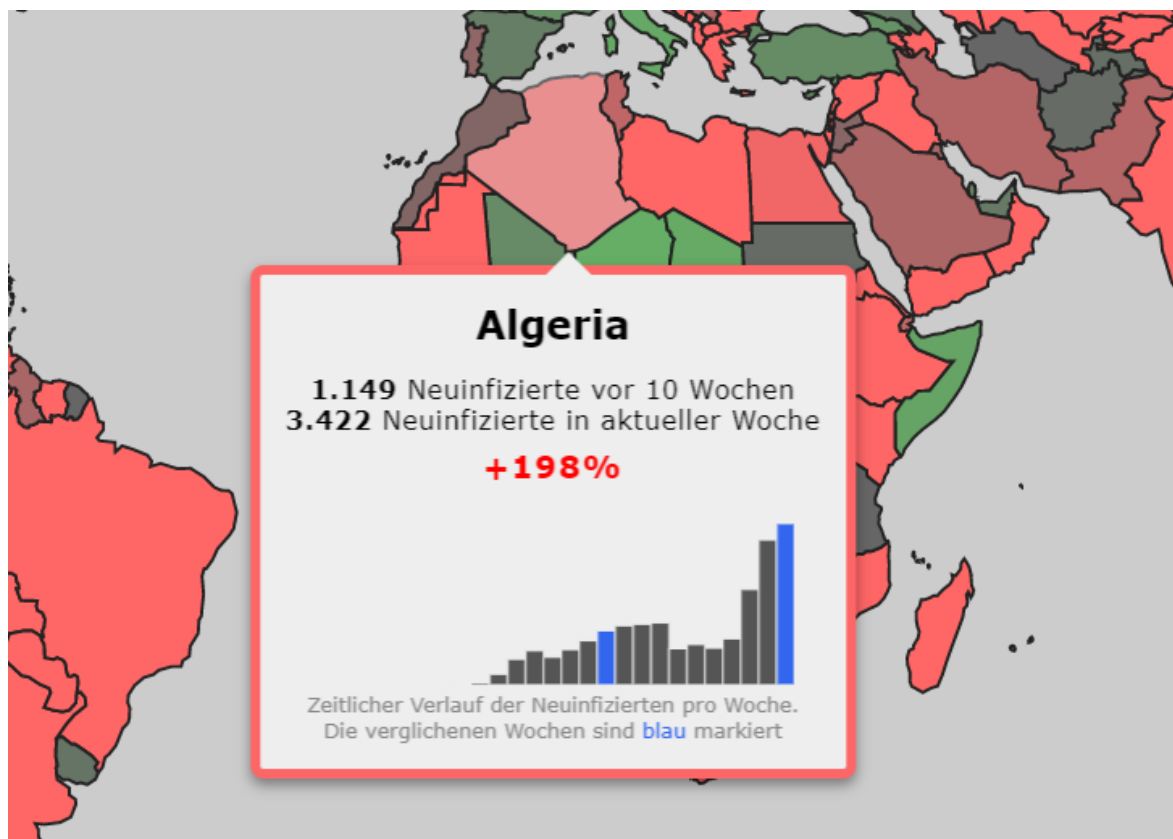
Durch das Visualisieren dieser relativen Unterschiede mit der Farbe des Landes lassen sich schnell Verbindungen zwischen Ländern erkennen. Zum Beispiel wenn alle Länder eines Kontinents die gleiche Farbe besitzen. Außerdem lässt sich auch aufgrund des hohen Kontrasts zwischen Rot und Grün ebenfalls schnell erkennen, wenn ein Land sich gegen den Trend anderer Länder entwickelt.

View Transformations

Der Benutzer hat die Möglichkeit, den Abstand zwischen den beiden zu vergleichenden Wochen mit dem Dialog unten links zu wählen. Dabei kann der Abstand einen beliebigen Wert über Null annehmen. Wobei ein Wert höher als 27 zum jetzigen Zeitpunkt keinen Sinn macht, da vor 28 Wochen noch keine Person mit Corona infiziert war. Eine Änderung dieser Zahl durch den Plus- oder Minusknopf bewirkt eine sofortige Änderung der Länderfarben abhängig vom neuen relativen Unterschied der Neuinfektionen zwischen den beiden Wochen.

Außerdem kann der Benutzer mehr Information zu einem Land erhalten, wenn er die Maus über ein Land bewegt. Dabei öffnet sich ein Dialog über oder unter dem Land, der den Namen des Landes, die Neuinfektionszahlen der vergangenen und der aktuellen Woche und den daraus resultierenden relativen Unterschied enthält.

Anschließend an diese Informationen werden in einem Balkendiagramm noch alle Wochen mit ihren jeweiligen Infektionszahlen dargestellt, um den Benutzer einen noch besseren Eindruck in die langfristige Entwicklung der Neuinfektionen in diesem Land zu geben. Dabei sind die beiden verglichenen Wochen mit blauer Farbe hervorgehoben.



Begründung des Satz von Tufte

Der Satz von Tufte besagt: „Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.“

Der Satz lässt sich damit interpretieren, dass eine grafische Darstellung als besonders gelungen gilt, wenn der Betrachter mit minimalem Aufwand aus dieser Darstellung neue Erkenntnisse gewinnen kann.

Meine Studienarbeit ist ein sehr gutes Beispiel für diesen Satz, da sie auf eine minimalistische Darstellung setzt, den Betrachter nicht mit zu vielen Konfigurationsmöglichkeiten verwirrt und trotzdem das Wichtigste an der Corona-Pandemie darstellt: Die Entwicklung der Neuinfektionen im Vergleich zu Vergangenheit.

Nur wenn Neuinfektionen abnehmen und ein Land das bei einem Virus übliche exponentielle Wachstum verlässt, kann es eine derartige Pandemie besiegen. Meine Visualisierung macht es dem Betrachter leicht, diese Entwicklung mithilfe des gewohnten Anblicks der Welt und den auffälligen Länderfarben zu erkennen. Und für noch mehr Informationen muss der Betrachter nur die Maus über dieses bewegen. Dann erhält er alle Werte, mit denen die Farbe des Landes bestimmt wird und durch das angefügte Balkendiagramm einen umfangreicheren in die Entwicklung der Neuinfektionen für dieses Land.