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. Da es eines der ersten Ergebnisse bei der Google-Suche nach „covid 19 data“ war und einen vielversprechenden Umfang besaß, war die Suche nach einem geeigneten Datenset schnell zu Ende. Es liegt als csv-Datei im Unterordner /data vor.

Zuerst wollte ich mir einen Überblick über den Umfang des Datenset machen, 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. Wohingegen Deutschland nach einem klar erkennbaren Anstieg der Neuinfektionen seitdem sich in einem stetigen Fall dieser Zahl befindet, scheint bei den USA nach dem ersten Anstieg ein zweiter noch stärkerer Anstieg zu folgen, ohne dass die Anzahl der Neuinfektionen dazwischen erkennbar abgenommen hätten. Dieser drastische Unterschied zwischen beiden Kurven führte zu der 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 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 Attributwerte enthielten. Wie zum Beispiel viele 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 Exports liegt wieder als csv-Datei im Unterordner /web/data vor.

# **Visual Mapping**

Nach erstem Einlesen mithilfe des Frameworks Panda wurde schnell klar, dass dieses Dataset viele Datenspalten enthält, die nicht für meine Visualisierung benötigt werden.

Welt von <https://simplemaps.com/resources/svg-world>

# **View Transformations**

Nach erstem Einlesen mithilfe des Frameworks Panda wurde schnell klar, dass dieses Dataset viele Datenspalten enthält, die nicht für meine Visualisierung benötigt werden.

# Begründung des Satz von Tufte

Nach erstem Einlesen mithilfe des Frameworks Panda wurde schnell klar, dass dieses Dataset viele Datenspalten enthält, die nicht für meine Visualisierung benötigt werden.