

---

## Philipp Dubach

Bogenstrasse 3  
5614 Sarmenstorf  
(+41) 79 610 99 22

# VISUALISIERUNGS-PROJEKT

4. September 2018 / CAS Datenvisualisierung FFHS Zürich

<b>SPEZIFIKATION DES VISUALISIERUNGS-PROJEKT</b>	<b>3</b>
WAS SOLL VISUALISIERT WERDEN?	3
WER IST DIE ZIELGRUPPE?	3
WAS SOLL DIE ZIELGRUPPE AUS DER VISUALISIERUNG ERKENNEN?	3
AUSWAHL DER WERKZEUGE FÜR DIE VISUALISIERUNG	4
AUSWAHL DER DATEN	4
ZUSAMMENFASSUNG DER SPEZIFIKATION	4
<b>SPEZIFIKATION DER DATEN</b>	<b>5</b>
DATENSATZ 1 - KAGGLE	5
QUELLE	5
INHALT / NUTZUNG	5
DATENSATZ 2 - DSLABS	5
QUELLE	5
INHALT / NUTZUNG	5
<b>ENTWURF</b>	<b>6</b>
VORANSICHT ENTWURF	6
<b>UMSETZUNG UND VALIDIERUNG</b>	<b>7</b>
UMSETZUNG	7
VALIDIERUNG	7
<b>SCHLUSSABGABE UND FEEDBACK</b>	<b>8</b>
PRÄSENTATION	8
HERLEITUNG DER DATEN	8
PERSÖNLICHES FAZIT	8

---

## Versionslog

Version	Datum	Inhalt
0.1	04.10.18	Abgabe Spezifikation der Visualisierungs-Projekt
0.2	28.10.18	Abgabe Spezifikation der Daten / Entwurf
0.3	06.12.18	Abgabe Umsetzung und Validierung
1.0	22.12.18	Schlussabgabe und Feedback

---

## **SPEZIFIKATION DES VISUALISIERUNGS-PROJEKT**

### **WAS SOLL VISUALISIERT WERDEN?**

Ich werde für dieses Visualisierungs-Projekt die Kriminalität der USA abbilden. Dies werde ich mit mehreren Visualisierungen darstellen und dazu eine Art Übersicht auf einer A4 Seite erstellen. Auf dieser Seite sind dann alle relevanten Visualisierungen sauber und schön zusammengefasst damit es druckbereit für zum Beispiel ein Magazin wäre. Die Visualisierungen sollen vor allem den zeitlichen Verlauf der Kriminalität zeigen und auch in welchen Regionen sich die höchste Kriminalität in den USA manifestiert. Wenn irgendwie möglich, möchte ich Geodaten dazu verwenden um die Kriminalitätsrate der USA zu kartographieren, damit auf den ersten Blick zu erkennen ist, wie sich die Gewalt in den USA verteilt. Ich möchte auch noch eine Visualisierung präsentieren, welche darstellt, ob die Bevölkerungsdichte mit der Mordrate harmonisierend ist.

### **WER IST DIE ZIELGRUPPE?**

Da sich die Themenwahl auf die USA begrenzt, ist es ein nationales Thema. Nichtsdestotrotz wird auf internationaler Ebene gerne auf die USA geblickt. Demnach ist die Zielgruppe etwas für Leute welche sich für die USA interessieren sowie für Personen welche sich für weltweite Themen interessieren welche wissen möchte, was in den anderen Nationen so läuft. Interessant könnte es auch für reisende Touristen oder Auswanderer sein. Ich denke sogar, dass das Interesse für dieses Thema für die meisten Personen recht interessant ist, weil Kriminalität und Gewalt quasi das Zugpferd von den meisten Journalisten und weltweiten Medien ist.

- Touristen / Reisebüros
- Auswanderer
- Interessierte an weltweiten Nachrichten

### **WAS SOLL DIE ZIELGRUPPE AUS DER VISUALISIERUNG ERKENNEN?**

Die Zielgruppe soll eine Übersicht über die Entwicklung der Kriminalität der USA erhalten. Sowie sollen sie erkennen können, wo in den USA es am sichersten und wo es gefährlichsten ist. Zudem soll auch aufgezeigt werden, ob es entscheidende Faktoren für die vorhandene Kriminalität gibt. Interessanterweise soll man feststellen können, ob die Kriminalität in hochbevölkerten Regionen höher ist als in kleinen Regionen ist. Eventuell kann die Visualisierung auch zur Auswahl eines sicheren Reiseziel in den USA fungieren.

---

## AUSWAHL DER WERKZEUGE FÜR DIE VISUALISIERUNG

Um die Daten zu bereinigen, berechnen und zu visualisieren werden ich R mit R Studio verwenden. Ich werde höchstwahrscheinlich die Library "ggplot2" verwenden. Die Abschlussarbeit für mein Titelblatt mit allen Visualisierungen, werde ich dann mit einer vektororientierten Grafiklösung fertigstellen, um der Arbeit den letzten Schliff zu verpassen.

## AUSWAHL DER DATEN

Als Datenquelle habe ich bereits ein paar Datensätze bei Kaggle.com gefunden, welche in die nähere Auswahl kommen. Diese kommen als CSV Datei daher. Zusätzlich habe ich noch einen weiteren Datensatz in der Library "dslabs" gesehen, welcher sich zusätzlich anbietet. Diese zwei Datensätze könnte ich dann mit Hilfe von R verknüpfen, damit noch mehr Analysen möglich sind. Bis jetzt habe ich mich jedoch noch nicht komplett auf diese zwei Datensätze festgelegt.

Die Daten müssen mindestens alle Gewalttaten, Morde, Verletzte, Zeitpunkt, Ort, Geodaten enthalten. Um das Ganze zu analysieren und zu Gruppieren ist es dann nötig alle Staaten, Regionen und die Bevölkerungsdichte zu kennen. Dafür sind mindestens zwei Datensätze nötig.

## ZUSAMMENFASSUNG DER SPEZIFIKATION

Das Projekt soll ein Bild darüber abgeben, wie es mit der Kriminalität in der USA steht. Dafür wird eine A4 Seite erstellt, welches dies auf den ersten Blick erkennen lässt. Zusätzlich soll dargestellt werden wo es in den USA viel Kriminalität gibt und wie sich die Gewalt über die Jahre verändert hat. Ziel ist die allgemeine Information für weltoffene Personen und Reisende/Auswanderer. Ausgeführt und analysiert wird die Aufgabe von Philipp Dubach.

---

## **SPEZIFIKATION DER DATEN**

Wie bereits erwähnt, werden ich zwei Datensätze verwenden. Ein Datensatz kommt von Kaggle und beschreibt alle Gewalttaten der USA. Der andere Datensatz gibt mir eine Liste der Bevölkerung. Ich werde diese zwei Datensatz nun genau definieren.

### **DATENSATZ 1 - KAGGLE**

#### **QUELLE**

[https://www.kaggle.com/jameslko/gun-violence-data#gun-violence-data\\_01-2013\\_03-2018.csv](https://www.kaggle.com/jameslko/gun-violence-data#gun-violence-data_01-2013_03-2018.csv)

#### **INHALT / NUTZUNG**

Alle gelisteten Gewalttaten der USA von 2014 - 2017 mit Angaben von Zeit, Ort, Opfer, Verletzte.

### **DATENSATZ 2 - DSLABS**

#### **QUELLE**

```
R-Studio Package  
install.packages("dslabs")  
library(dslabs)  
dslabs::murders
```

#### **INHALT / NUTZUNG**

Alle Staaten der USA mit Angaben von Bevölkerung, Regionen, Morde

## ENTWURF

Ein erster Entwurf wurde erstellt

- Das Geo-Mapping wurden mit R Studio in Kombination mit “ggplot” erstellt.
- Das Verlaufsdiagramm der Kriminalität über die Jahre mit “Google Tabellen” anhand einem CSV-Export von R Studio.
- Das Regressionsdiagramm mit R Studio und “ggplot”

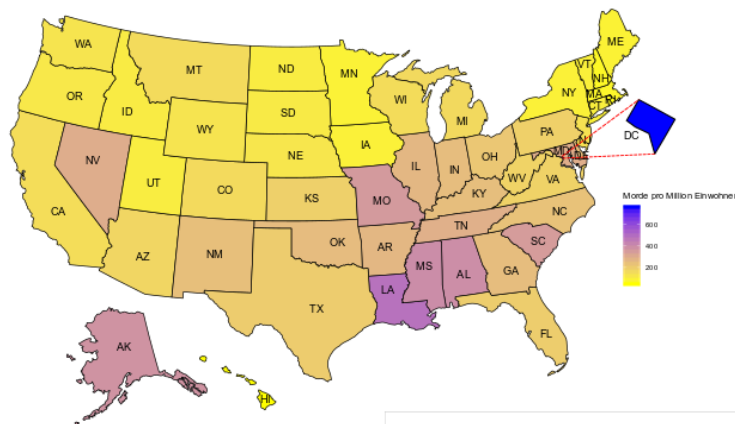
## VORANSICHT ENTWURF

### Die Kriminalität der USA

ENTWURF

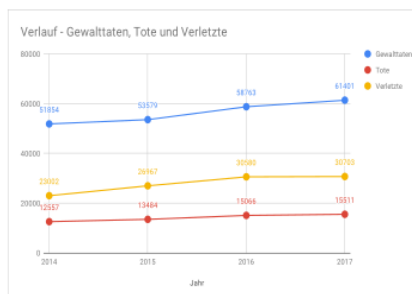
Der original Entwurf wird als separates PDF zur Verfügung gestellt.

Anzahl Morde pro Million Einwohner  
Zeitraum 2014 – 2017 / PD 2015

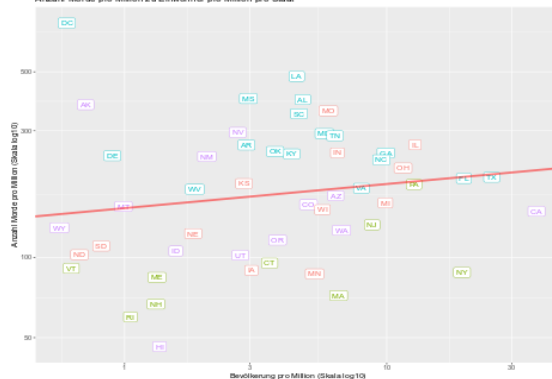


Shortfacts:  
Durchschnittlich gibt es alle 11,5 Minuten eine Gewalttat.  
Bei 50% wird jemand Verletzt und in 25% der Fällen gibt es Tote.

Shortfacts:  
Tendenziell steigt die Kriminalität mit den letzten Jahren leicht an.



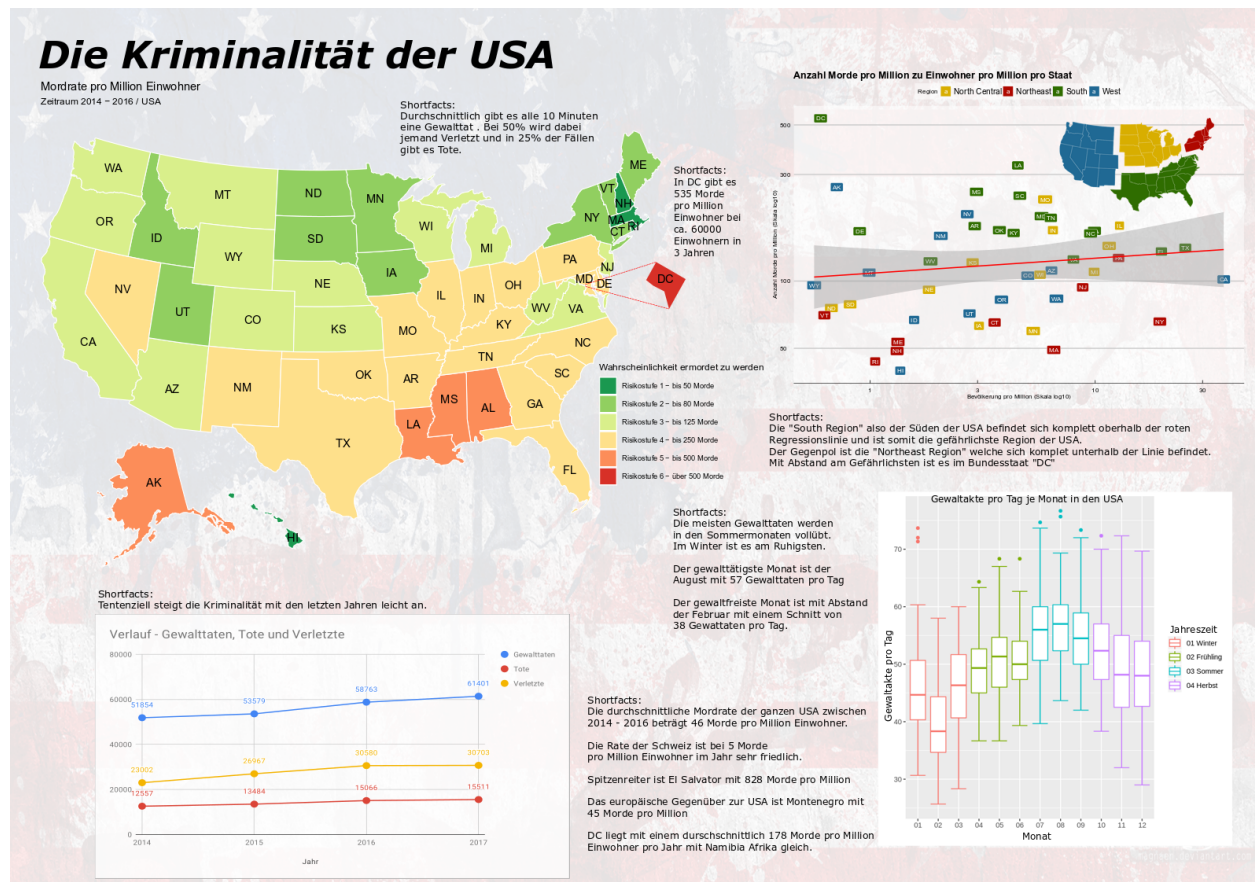
Anzahl Morde pro Million zu Einwohner pro Million pro Staat



Shortfacts:  
Die "South Region" also der Süden der USA befindet sich komplett oberhalb der roten Regressionslinie und ist somit die gefährlichste Region der USA. Der Gegenpol ist die "Northwest Region" welche sich komplett unterhalb der Linie befindet. Mit Abstand am Gefährlichsten ist es im Bundesstaat "DC"

# UMSETZUNG UND VALIDIERUNG

## UMSETZUNG



Original liegt als PDF im Format A3 bei

## VALIDIERUNG

Meine Daten habe ich anhand rechenen im Internet geprüft. Die Daten stimmen so plus/minus 10%. Fakt war, dass es sehr viele unterschiedliche Quellen für die Kriminalität in den USA gibt mit jeweils unterschiedlichen Aspekten. Dazu folgt noch, dass die Einwohnerzahl der Staaten und der ganzen USA teilweise auch komplett unterschiedlich sind. Gemäss diversen Wikipedia Berichten, sollten meine Daten der Wirklichkeit nahe kommen.

Folgender Wikipedia Link:

[https://de.wikipedia.org/wiki/T%C3%B6tungsrate\\_nach\\_L%C3%A4ndern](https://de.wikipedia.org/wiki/T%C3%B6tungsrate_nach_L%C3%A4ndern)

[https://en.wikipedia.org/wiki/Gun\\_violence\\_in\\_the\\_United\\_States](https://en.wikipedia.org/wiki/Gun_violence_in_the_United_States)

[https://en.wikipedia.org/wiki/Crime\\_in\\_the\\_United\\_States](https://en.wikipedia.org/wiki/Crime_in_the_United_States)

---

## SCHLUSSABGABE UND FEEDBACK

### PRÄSENTATION

Die Präsentation wurde mit Prezi erstellt und kann mit folgendem Link angesehen werden:

[http://prezi.com/i8ilyb97ufg5/?utm\\_campaign=share&utm\\_medium=copy](http://prezi.com/i8ilyb97ufg5/?utm_campaign=share&utm_medium=copy)

### HERLEITUNG DER DATEN

Alle Scripts mit der Herleitung sowie alle Dateien habe ich bei GitHub abgelegt unter folgendem Link:

<https://github.com/dubachphil/datavisualisation>

### PERSÖNLICHES FAZIT

Ich bin eigentlich ziemlich zufrieden mit der Arbeit. Ich hätte unter Umständen lieber noch etwas Dynamisches gemacht, jedoch wollte ich nicht von meiner anfänglichen Spezifikation abweichen. Es wäre schön gewesen, wenn ich für meine Präsentation die nötige Zeit zum Präsentieren bekommen hätte. Leider war die Zeit ziemlich rar gesät und ich musste die Präsentation im Schnelldurchlauf präsentieren. Ich kann auf jeden Fall behaupten, dass ich etwas aus diesem Kurs mitnehme. Ich habe zuvor noch nie etwas mit Geo-Mapping gemacht und finde es sehr interessant, was alles damit visualisiert werden kann. Tools dafür gibt es ja im Übermaß, weshalb ich mich dann wirklich entschieden habe, es mit R zu machen. Einfach nur Daten einer Achse zuweisen wollte ich dann schon nicht.

Besten Dank für den Kurs