

Semesterarbeit Explorative Datenanalyse

Gamper Gian, Bratschi Jonas

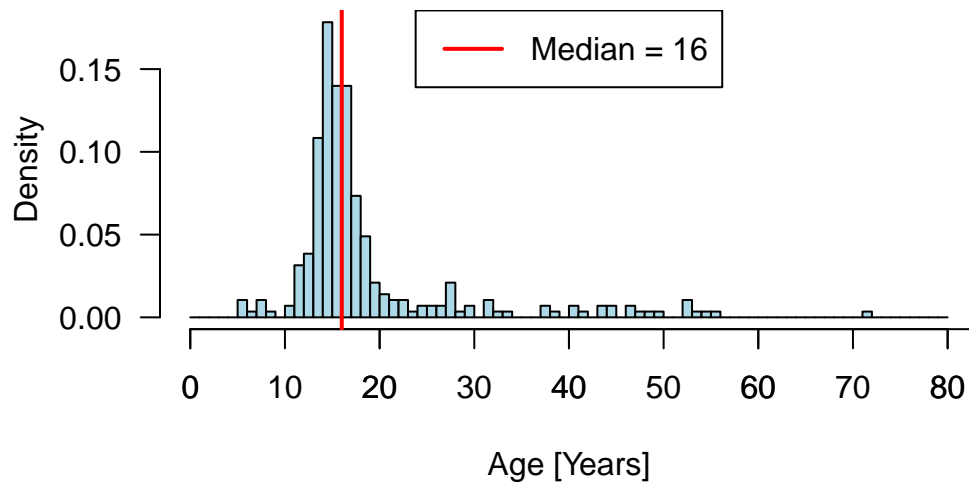
Einleitung

In dieser Semesterarbeit analysieren wir Daten zu Schießereien der letzten drei Jahrzehnte an US-Amerikanischen Schulen. Der Datensatz umfasst, basierend auf Forschung von Pah/Amaral/Hagan und Wikipedia, eine Bandbreite von Vorfällen, von unbeabsichtigten Schüssen bis zu Massenmorden. Im Fokus steht die relative Gefährdung von Schülern durch Schusswaffen, unabhängig von den Tätermotiven. Den Datensatz stammt von der Website Kaggle (<https://www.kaggle.com/code/ecodan/us-school-shootings-dataset>).

Tabelle mit den Variablen und deren beschreibung

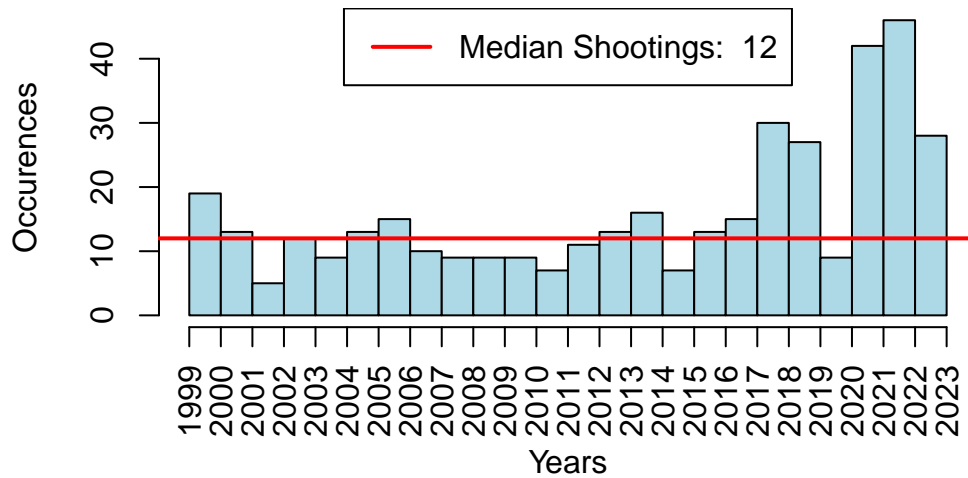
Variable	Beschreibung
uid	Einzigartige Identifikationsnummer
nces_school_id	NCES Schul-ID
school_name	Name der Schule
nces_district_id	NCES Distrikt-ID
district_name	Name des Schuldistrikts
date	Datum des Ereignisses
school_year	Schuljahr des Ereignisses
year	Jahr des Ereignisses
time	Uhrzeit des Ereignisses
day_of_week	Wochentag des Ereignisses
city	Stadt der Schule
state	Bundesstaat der Schule
school_type	Schultyp
enrollment	Anzahl der eingeschriebenen Schüler
killed	Anzahl der Todesopfer
injured	Anzahl der Verletzten
casualties	Gesamtzahl der Opfer
shooting_type	Art des Schießereignisses
age_shooter1	Alter des ersten Schützen
gender_shooter1	Geschlecht des ersten Schützen
race_ethnicity_shooter1	Rasse/Ethnizität des ersten Schützen
shooter_relationship1	Beziehung des ersten Schützen zur Schule
shooter_deceased1	Informationen, ob der erste Schütze verstorben ist
deceased_notes1	Anmerkungen zum Tod des ersten Schützen
age_shooter2	Alter des zweiten Schützen
gender_shooter2	Geschlecht des zweiten Schützen
race_ethnicity_shooter2	Rasse/Ethnizität des zweiten Schützen
shooter_relationship2	Beziehung des zweiten Schützen zur Schule
shooter_deceased2	Informationen, ob der zweite Schütze verstorben ist
deceased_notes2	Anmerkungen zum Tod des zweiten Schützen

Age of school shooters



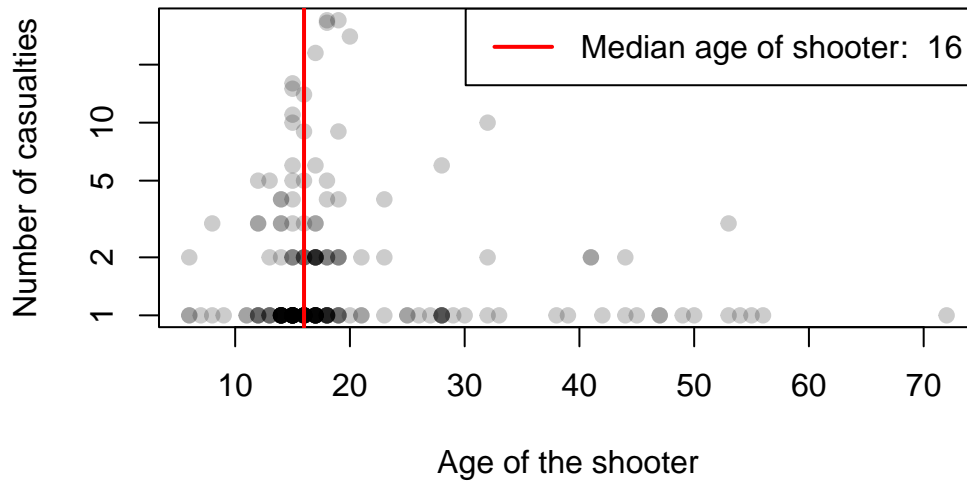
Die Grafik zeigt ein Histogramm, das das Alter von Schützen auf der X-Achse mit dem Medianwert bei 16 Jahren und der relativen Häufigkeit auf der Y-Achse darstellt. Primär konzentriert sich die Verteilung auf das Teenager Alter, was auf Schüler:innen hindeutet. Jedoch gibt es auch einige Schützen, welche ein Alter deutlich über dem typischen Schulalter aufweisen. Dies könnten möglicherweise Lehrer:innen oder andere Angestellte der Schule sein.

Numbers of school shootings in the US since 1999

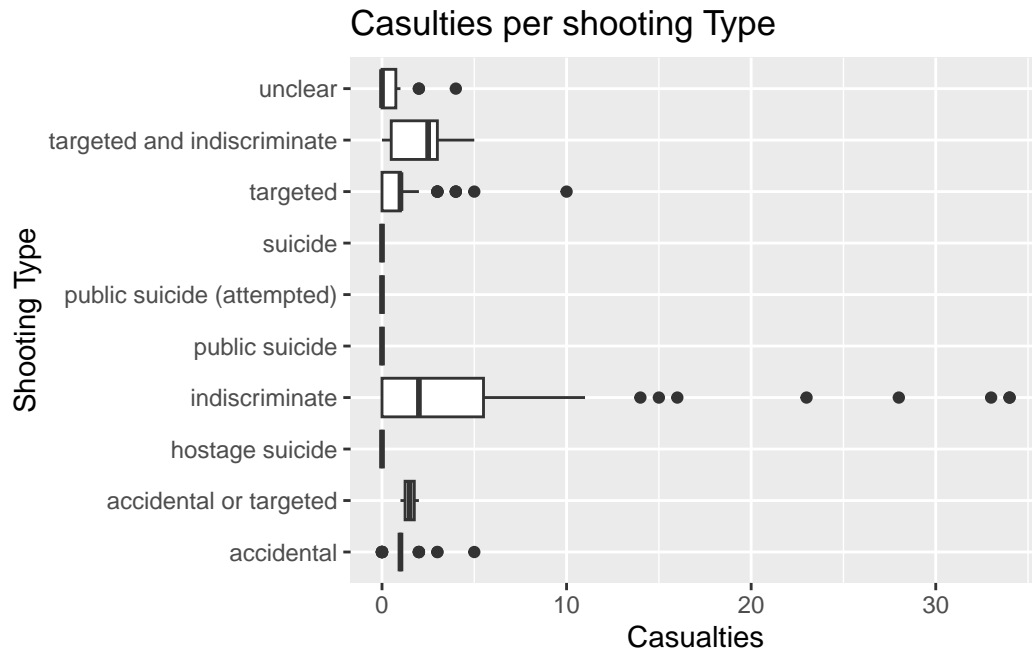


Dieses Histogramm visualisiert die Anzahl an Schiessereien an Schulen für jedes Jahr, seit 1999. Der Median seit 1999 liegt hierbei bei 12 vorfällen pro Jahr. Über die Aussage des Medians in dieser Grafik lässt sich streiten. Nichtsdestotrotz ist aber eine klare Zunahme der Fälle seit 2017 zu erkennen.

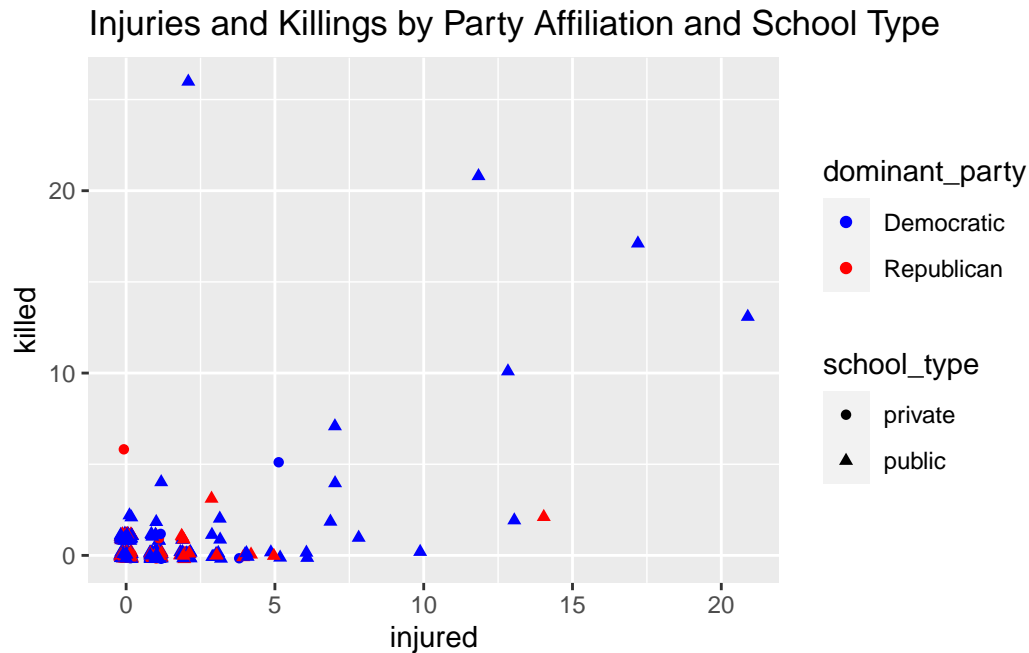
Shooter Age / Number of casualties



In diesem Scatterplot sind auf der X-Achse die Altersangaben der Schützen und auf der Y-Achse die Anzahl der “Casualties” dargestellt, also der Toten und Verwundeten. Der Median des Schützenalters liegt bei 16 Jahren. Deutlich wird in der Grafik, dass diejenigen, die erheblichen Schaden verursacht haben, größtenteils im Altersbereich von 14 bis 20 Jahren anzutreffen sind. Dies legt nahe, dass diese Schützen möglicherweise klare Absichten hatten und entschlossen handelten, wobei wir annehmen, dass das Tatmotiv in dieser Altersgruppe wahrscheinlich auf einen Amoklauf zurückzuführen ist. Andererseits fällt auf, dass es sowohl sehr junge als auch sehr alte Schützen gibt, bei denen wir davon ausgehen, dass es sich hauptsächlich um Unfälle handelt.



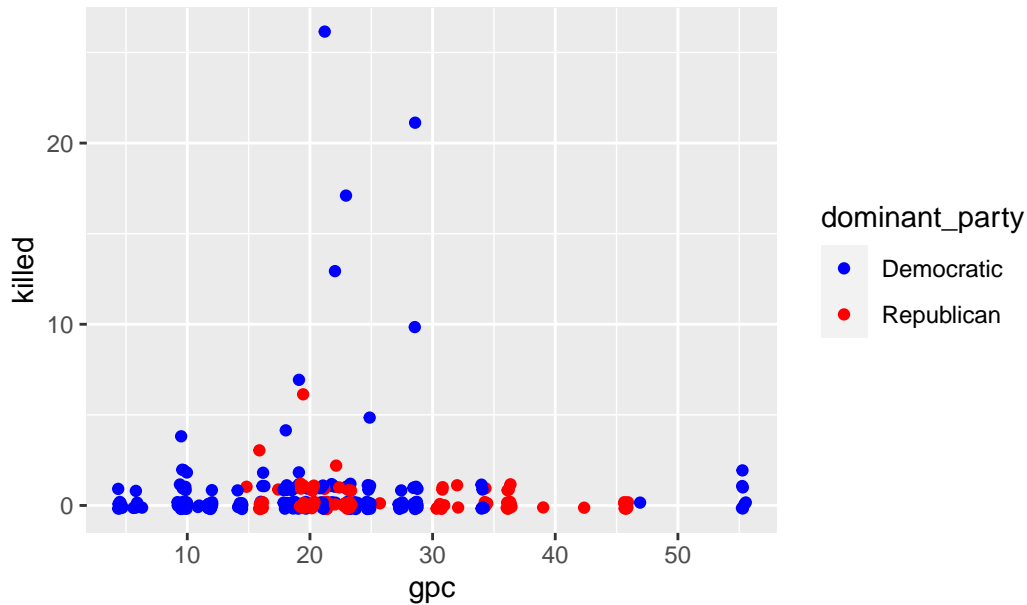
Die Obenstehende Boxplot Grafik stellt die Anzahl Opfer nach Art des Vorfalls dar. Dabei ist klar zu erkennen, dass bei Indiscriminate die grösste Streuung vorhanden ist. Dies deutet darauf hin, dass die Opferzahlen bei solchen Vorfällen sehr unterschiedlich ausfallen. Bei Suiziden sowie Unfällen sind die Opferzahlen erwartungsgemäss eher gering.



In diesem Diagramm sind auf der X-Achse die Anzahl der verletzten Personen und auf der Y-Achse die Anzahl der getöteten Personen pro Vorfall dargestellt. Zusätzlich haben wir unseren Datensatz durch Informationen zur politischen Ausrichtung jedes Bundesstaates erweitert, die aus dem Wahljahr 2020 stammen. Die Symbole der Datenpunkte geben zudem Auskunft darüber, ob es sich um eine öffentliche oder private Schule handelt. Bemerkenswert ist, dass Schießereien anscheinend überwiegend an öffentlichen Schulen auftreten. Des Weiteren fällt auf, dass Schießereien mit vielen Toten und Verletzten hauptsächlich in demokratischen Staaten stattzufinden scheinen. Wir gehen davon aus, dass diese Korrelation auf eine Vielzahl von Faktoren zurückzuführen ist. Hierbei könnten die Hauptfaktoren unter anderem die Strenge der Waffengesetze, der allgemeine Zugang zu Waffen, sozioökonomische Unterschiede sowie die generelle politische Lage in den Bundesstaaten sein.

Wichtig ist aber, dass man anhand von dieser Grafik keine Kausalität ableitet. Und auch das Simpsonsche Paradoxon ist hier nicht zu vernachlässigen, könnte es zum Beispiel sein, dass in demokratischen Staaten mehr Schulen sind oder allgemein die Bevölkerung jünger ist und somit mehr Personen zur Schule gehen?

Gun Ownership per Capita vs. Killings by Party Affiliation



In diesem Diagramm sind auf der X-Achse die Waffen pro 100 Einwohner (Guns per Capita) für den jeweiligen Bundesstaat gemäss einem Datensatz aus dem Jahr 2023 und auf der Y-Achse die Anzahl der getöteten Personen pro Vorfall dargestellt. Zusätzlich haben wir unseren Datensatz durch Informationen zur politischen Ausrichtung jedes Bundesstaates erweitert, die aus dem Wahljahr 2020 stammen. Hier ist zu erkennen, dass die meisten Vorfälle in Bundesstaaten mit einem gpc von 10 bis 30 stattfanden. Ein Anstieg von vorfällen in Bundesstaaten mit höheren GPC ist aufgrund dieser Daten nicht zu Beobachten.