

Morde in Chicago

Fallstudie Applied Analytics | WS 22/23

Der Datensatz

Die Stadt Chicago stellt Informationen zu aufgenommenen Straftaten frei online zur Verfügung. Der Datensatz enthält zu jeder der mehr als 7,7 Millionen enthaltenen Straftaten rudimentäre Informationen wie etwa die Art der Straftat, das Datum und die Uhrzeit der Aufnahme durch die Polizei, den Ort oder ob eine Festnahme erfolgt ist oder nicht. Die Daten reichen bis in das Jahr 2001 zurück und werden täglich aktualisiert. Für die Fallstudie liegen die Daten bis einschließlich dem 30.12.2022 vor. Um euch den Start zu erleichtern, habe ich die Daten bereits geringfügig transformiert:

- Duplikate sind entfernt
- Falsch erkannte Datentypen sind korrigiert, insbesondere Datumsfelder
- Unnötige Felder wie `case_id` sind entfernt, um die Größe der Daten um fast ein Drittel zu reduzieren
- Die Daten liegen im R-internen Format `.rds` vor

Erstellt ein neues Projekt für die Bearbeitung der Fallstudie. [Ladet den Datensatz herunter](#) und speichert ihn im Ordner `data/`. Anschließend könnt ihr die Daten wie folgt laden:

```
crimes <- readRDS("data/chicago_crimes_2001_to_2022.rds")
```

Überprüft anschließend die Anzahl Zeilen:

```
crimes %>% count()

# # A tibble: 1 x 1
#       n
#   <int>
# 1 7702153
```

Eure Aufgabe

Für die Lösung der Fallstudie sollt ihr einen Bericht mit Quarto verfassen, in dem ihr die vorliegenden Daten *explorativ* mit R und dem Tidyverse analysiert. Das Leitthema der explorativen Datenanalyse soll *die Entwicklung der Tötungsdelikte (homicide) über den verfügbaren Zeitraum* sein. Dabei sollt ihr in 3 Teilen zunächst einige Fakten zu den Straftaten ermitteln und diese Fakten visuell möglichst gut verständlich darstellen und erläutern (Teil 1). Im zweiten Teil geht es um die explorative Analyse zu offenen Fragestellungen, für die ihr eigene Ideen einbringen müsst und im Gegensatz zu Teil 1 kein klarer Weg vorgegeben ist. In Teil 3 geht es dann speziell um die visuelle Beantwortung von Fragen mithilfe von Kartendarstellungen.

Teil 1: Geschlossene Fragen

Stellt im ersten Teil eures Berichts einige relevante Fakten zu den Daten vor. Insbesondere die folgenden Fragen sollen mit einer geeigneten Visualisierung und einer präzisen textuellen Interpretation beantwortet werden:

1. Was sind die häufigsten begangenen Arten von Straftaten in 2022 und wo ordnen sich Mordfälle in dieser Liste ein?
2. Wie viele Mordfälle gab es in dem gesamten Zeitraum des Datensatzes pro Jahr?
3. Welche Straftat hat die niedrigste Festnahmerate, welche die höchste?
4. Wie ist die Festnahmerate bei Mordfällen insgesamt und in jedem Jahr?
5. Wie ist die Häufigkeit der Mordfälle prozentual im Durchschnitt über die Monate im gesamten Zeitraum verteilt?
6. Welche Polizeireviere (*police districts*) haben prozentual die meisten Mordfälle in den Abend- und Nachstunden zu verzeichnen und sollten vielleicht deshalb die Nachtschicht verstärken?
7. Welche Top 10 Straßen in Chicago haben eine besonders hohe Anzahl an Straftaten zu verzeichnen?

Teil 2: Offene Fragen

Im zweiten Teil des Berichts geht es um offene Fragen und eure Ideen, wie man diese auf interessante Weise analysieren und beantworten könnte:

1. Erstellt einen Vergleich zwischen der Entwicklung unterschiedlicher Arten von Schwerverbrechen mit denen von Tötungsdelikten (*homicide*).

2. Welche *Community Areas* haben im Verlauf der letzten 10 Jahre eine besonders positive Entwicklung bei der Bekämpfung schwerer Straftaten genommen?
3. Spielen sich mehr Mordfälle in Gebäuden oder im Freien ab? Gibt es saisonale Unterschiede?
4. Wie ist die Verteilung der Zeitabstände zwischen Mordfällen in den verschiedenen *Community Areas* im Vergleich? Gibt es hier Auffälligkeiten?
5. Zu welchen Uhrzeiten sollte man an welchen Orten besonders vorsichtig sein?

Teil 3: Kartendarstellungen

Erstellt für die folgenden Fragen aussagekräftige Kartendarstellungen und interpretiert das Ergebnis:

1. In welchen *Community Areas* der Stadt Chicago wurden in 2022 die meisten Mordfälle aufgenommen? Stellt die Analyse auch in den Zusammenhang zur [Einwohnerzahl der Community Areas](#)!
2. Gibt es Hotspots in Chicago für Straftaten im Zusammenhang mit Prostitution?

Bonusfragen

1. In der Wissenschaft gibt es Untersuchungen zu der Frage, ob es einen Zusammenhang zwischen Gewaltverbrechen und dem Wetter gibt. Konkret wurde ein positiver Zusammenhang zwischen Hitze und einer größeren Gewaltbereitschaft postuliert. Untersucht diesen Zusammenhang im vorliegenden Datensatz!
2. Sucht in den Daten nach besonders auffälligen Tagen und recherchiert, was zu der Auffälligkeit geführt hat!

Links

- Offizielle Webseite des Datensatzes auf dem [Chicago Data Portal](#) inklusive eines Data Dictionary.
- [Wikipedia-Seite der Chicago Community Areas](#) inklusive Angaben zur Einwohnerzahl und Fläche.
- [Boundaries der Chicago Community Areas](#), inklusive [Download als Shapefile](#).