

TU DORTMUND

FALLSTUDIEN I - SOMMERSEMESTER 2025

# **Projekt 1: Deskriptive Analyse von Ausgaben zur Kriminalitätsbekämpfung**

Dozenten:

JProf. Dr. Dennis Dobler

Loreen Sabel

Dr. Susanne Frick

Prof. Dr. Philipp Doeblen

Verfassener: Quang Vinh Nguyen

Gruppenmitglieder:

Alain Kauth, Ibrahim Rahman

22. April 2025

# Inhaltsverzeichnis

|  |           |
|--|-----------|
| <b>1. Einleitung</b>                                 | <b>1</b>  |
| <b>2. Problemstellung</b>                            | <b>1</b>  |
| 2.1. Datenmaterial . . . . .                         | 1         |
| 2.2. Ziele des Projekts . . . . .                    | 2         |
| <b>3. Statistische Methoden</b>                      | <b>3</b>  |
| 3.1. Univariate Verfahren . . . . .                  | 3         |
| 3.2. Multivariate Verfahren . . . . .                | 6         |
| <b>4. Statistische Auswertung</b>                    | <b>7</b>  |
| 4.1. Häufigkeitsverteilungen der Variablen . . . . . | 7         |
| 4.2. Zusammenhänge zwischen den Variablen . . . . .  | 8         |
| <b>5. Zusammenfassung</b>                            | <b>9</b>  |
| <b>Literatur</b>                                     | <b>10</b> |
| <b>A. Zusätzliche Grafiken</b>                       | <b>11</b> |

# 1. Einleitung

Die Bekämpfung der Kriminalität ist eine wichtige Aufgabe zur Wahrung der sozialen Sicherheit. Allerdings erfordert diese Aufgabe zweifellos erhebliche Investitionen in das Strafrechtssystem, einschließlich der Polizei, der öffentlichen Verwaltung und der Gerichte. Die Vereinigten Staaten sind seit langem eines der einflussreichsten Länder der Welt. Demnach finden die US-Ausgaben für die Kriminalprävention auch große Beachtung.

Im Rahmen dieses Projekts werden die Ausgaben der US-Bundesstaaten für die Verbrechensbekämpfung sowie damit verbundene Variablen untersucht. Im Wesentlichen fokussiert sich das Projekt auf univariate als auch multivariate Verfahren der deskriptiven Analyse, um potenzielle Faktoren zu identifizieren, die die Ausgaben der US-Bundesstaaten beeinflussen.

Der vorliegende Bericht ist in fünf Kapitel unterteilt. In Kapitel 2 erfolgt die Präsentation des Datensatzes über US-staatliche Ausgaben für die Kriminalitätskontrolle. In Kapitel 3 werden grundlegende deskriptive und explorative Methoden zur Analyse vorgestellt. Dabei wird auf wichtige Kennzahlen wie Mittelwert, Median, Korrelationskoeffizient als auch grafische Methoden wie Histogramm, Scatterplot eingegangen. Kapitel 4 befasst sich mit den möglichen Zusammenhängen zwischen den Variablen im Datensatz. Von besonderem Interesse sind die Zusammenhänge mit den Ausgaben. Zusätzlich wird auch diskutiert, welche der Variablen einen Einfluss auf die Werte der anderen haben können. Abschließend werden die wichtigsten Ergebnisse in Kapitel 5 zusammengefasst.

## 2. Problemstellung

In diesem Kapitel werden die Projektziele sowie das Datenmaterial ausführlich dargelegt.

### 2.1. Datenmaterial

Der verwendete Datensatz wurde im Rahmen der Veranstaltung von den Dozierenden bereit gestellt und umfasst Informationen zur Verbrechensbekämpfung und den entsprechenden Ausgaben aus den 50 US-Bundesstaaten und DC. Somit gibt es 51 Einträge, die jeweils aus acht Variablen bestehen. Die in die Untersuchung einbezogenen Variablen werden wie folgt beschrieben:

- **Staat:** (nominal/{AK, AL, ..., WY}) Die Abkürzung des US-Bundesstaates;
- **Ausgaben:** (metrisch/positive natürliche Zahlen) Staatsausgaben (in \$1.000) im Zusammenhang mit kriminellen Aktivitäten;
- **Beobachtung:** (metrisch/positive Zahlen) Die Anzahl Personen unter kriminalistischer Beobachtung;
- **Verbrechensrate:** (metrisch/positive natürliche Zahlen) Anzahl Verbrechen pro 100.000 Einwohner;
- **Anwälte:** (metrisch/positive natürliche Zahlen) Anzahl Anwälte im Staat;
- **Staatsdiener:** (metrisch/positive natürliche Zahlen) Anzahl Personen, die einer Erwerbstätigkeit für den Staat nachgehen und für diesen eine Dienstleistung ausführen;
- **Einwohner:** (metrisch/positive natürliche Zahlen) Anzahl Einwohner im Staat (in 1.000);
- **Fläche:** (metrisch/positive Zahlen) Gesamtfläche des Staats (in 1.000 km<sup>2</sup> ), also Land- und Wasserfläche kombiniert;
- **Konföderation:** (dichotom/{ja, nein}) Angaben, ob ein Staat in den 1860er-Jahren Teil der Konföderierten Staaten von Amerika waren bzw. von diesen beansprucht wurde .

Es liegen keine weiterführenden Informationen bezüglich der Herkunft oder Erhebung der Daten vor, sodass eine Beurteilung der Zuverlässigkeit, Genauigkeit usw. nicht möglich ist. Der Datensatz weist keine fehlenden Werte auf. Allerdings ist es als fragwürdig zu erachten, dass die Anzahl Personen unter kriminalistischer Beobachtung nicht eine natürliche Zahl, sondern eine Dezimalzahl ist. Des Weiteren sind im Datensatz keine Zeitinformationen enthalten, sodass die Aktualität der Studie nicht garantiert werden kann.

## 2.2. Ziele des Projekts

In erster Linie sind die Zusammenhänge zwischen den Variablen von Interesse. In Ergänzung dazu werden sie im Hinblick auf Ursache und Wirkung diskutiert. Auch für jede einzelne Variable wird ihre Stichprobe bzw. das Auftreten von Ausreißern ("außergewöhnlichen" Punkten) einer eingehenden Betrachtung unterzogen.

Im Rahmen der statistischen Zielsetzung liegt der Schwerpunkt des Projektvorhabens auf der deskriptiven Auswertung der Daten. Hierbei werden die Kennzahlen, das Histogramm und der Boxplot als univariate Verfahren eingesetzt, um die Stichprobe einzelner Variablen zu beschreiben und Ausreißer zu identifizieren. Beim multivariaten Verfahren ermöglicht die Kombination von Streudiagrammen bzw. Korrelationskoeffizienten die Analyse und Verallgemeinerung der Korrelation zwischen den Variablen.

## 3. Statistische Methoden

Das vorliegende Kapitel ist in zwei Abschnitte unterteilt: univariate und multivariate Verfahren. Die Datenverarbeitung und die Anwendung der hier aufgeführten Methoden erfolgt unter Verwendung der Software R mit Version 4.4.2 (Team, 2024) sowie des Pakets ggplot2 (Wickham et al., 2016).

### 3.1. Univariate Verfahren

Univariate Verfahren umfassen statistische Methoden für eindimensionale Daten (auch univariate Daten genannt). Dabei werden nur Beobachtungen eines einzigen Merkmals untersucht. Es wird von einer Stichprobe vom Umfang  $n$  ausgegangen, bei der die Werte  $x_1, \dots, x_n$  die beobachteten Werte einer Variablen  $X$  sind. Wenn die Skala der Variable  $X$  mindestens ordinal ist, lässt sich die Stichprobe auch als aufsteigende Reihenfolge  $x_{(1)} \leq \dots \leq x_{(i)} \leq \dots \leq x_{(n)}$  dargestellt werden. Dabei wird  $x_{(1)}$  als Minimalwert (Min) und  $x_{(n)}$  als Maximalwert (Max) bezeichnet.

Bevor auf die Beschreibung eingegangen wird, ist es notwendig, das Konzept von Lagemaß und von Streuungsmaß zu klären. Lagemaße sind die Kennzahlen, die das Zentrum einer Stichprobe darstellen. Streuungsmaße messen die Streuung der Daten um dieses Zentrum.

**Arithmetisches Mittel:** Das arithmetische Mittel ist ein wichtiges Lagemaß für metrische Variablen in Statistik. Man erhält den (arithmetische) Mittelwert, indem man die Summe der beobachteten Werte durch den Umfang der Stichprobe dividiert, also  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  (Fahrmeier et al., 1987, S.53-54). Zu berücksichtigen ist, dass das arithmetische Mittel empfindlich auf Extremwerte reagiert.

**Standardabweichung der Stichprobe:** Die Standardabweichung (eng. Standard deviation – SD)  $\sqrt{s^2}$  ist eine Maßzahl für die Streuung der Daten um den Mittelwert  $\bar{x}$ . Sie entspricht der Quadratwurzel der zugehörigen Stichprobenvarianz  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  (Fahrmeier et al., 1987, S.69-70), hat somit dieselbe Einheit wie die beobachteten Werte in der Stichprobe und erleichtert die Interpretation. Wie beim arithmetischen Mittel wird dabei eine metrische Variable für eine sinnvolle Anwendung vorausgesetzt. Dieses Streuungsmaß ist auch nicht resistent gegenüber Extremwerten.

**Quantile:** Das Konzept des Quantils bildet die Grundlage für die nachfolgenden Methoden. Die Anwendung bedingt die Anordnung der Stichprobe in aufsteigender Reihenfolge. Sei die Variable zudem metrisch, wird der Wert von p-Quantil,  $0 < p < 1$ , wie folgt bestimmt (Fahrmeier et al., 1987, S.64-65):

$$x_p = \begin{cases} x_{(\lceil np \rceil)} & \text{für } np \text{ nicht ganzzahlig,} \\ \frac{1}{2}(x_{(np)} + x_{(np+1)}) & \text{für } np \text{ ganzzahlig.} \end{cases}$$

Es gibt mindestens  $np$  Daten, die kleiner oder gleich  $x_p$  sind, sowie mindestens  $n - np$  Daten, die größer oder gleich  $x_p$  sind.

**Median:** Der Median dient als ein robustes Lagemaß. Unter dem Wort *robust* bzw. *resistent* wird verstanden, dass der Einfluss extremer Werte auf eine Methode begrenzt wird. Das gilt,  $x_{med} = x_{0.5}$  (Fahrmeier et al., 1987, S.55-56,64-65). Der Median befindet sich in der Datenmitte und teilt die Datenmenge in zwei Teile auf. Sowohl der obere als auch der untere Teil umfassen mindestens die Hälfte der Daten. Die Positionen der anderen Daten in den beiden Hälften haben keinen Einfluss auf den Median. Aus diesem Grund weist der Median eine höhere Robustheit im Vergleich zum arithmetischen Mittel auf.

**Interquartilsabstand:** Die obere bzw. untere Hälfte kann man wiederum durch einen Wert halbieren. Dieser Wert wird als der obere Quartil  $Q_3$  bzw. untere Quartil  $Q_1$  bezeichnet und entspricht jeweils  $x_{0.75}$  bzw.  $x_{0.25}$ . Aus beiden Quartilen lässt sich ein robuste Streuungsmaß ableiten. Dieses Streuungsmaß  $d_Q = Q_3 - Q_1$  wird als der Interquartilsabstand (eng. Interquartile range – IQR) bezeichnet (Fahrmeier et al., 1987, S.66). Etwa die Hälfte der Daten fällt in den IQR. Ähnlich wie beim Median haben die Positionen der Daten links vom  $Q_1$  und rechts vom  $Q_3$  auch keinen Einfluss auf den IQR. Dementsprechend gilt IQR als robust.

**Boxplot:** Der Boxplot kann eine kompakte visuelle Zusammenfassung der Stichprobe einer metrischen Variable bieten. Der Boxplot besteht aus einer Box, einer Linie innerhalb der Box, welche den Median der Daten anzeigt, sowie zwei Whiskers und gegebenenfalls Markierungen für alle Ausreißer. Die zwei Ränder der Box stellen jeweils den oberen und unteren Quartil dar. Somit stimmt die Länge der Box mit dem IQR überein.

Die weiteren Beschreibungen basieren auf der modifizierten Boxplot-Version nach Tukey (ggplot2, 2024), da diese im Rahmen des Projekts zum Einsatz kommt. Der obere Whisker erstreckt sich vom oberen Rand bis zum größten Wert, der nicht weiter als  $1.5 \cdot \text{IQR}$  vom oberen Rand entfernt ist. Der untere Whisker erstreckt sich analog vom unteren Rand bis zum kleinsten Wert innerhalb des Bereichs von  $1.5 \cdot \text{IQR}$ . Daten über das Ende der Whiskers hinaus werden als Ausreißer bezeichnet und einzeln markiert.

**Histogramm:** Das Histogramm ist ein Visualisierungsmittel für metrische Variablen (Fahrmeier et al., 1987, S. 41-42). Angenommen, dass die Variable  $X$  den Ausprägungsbereich  $[c_0, c_k)$  aufweist, kann dieser in  $k$  benachbarte Intervalle  $[c_0, c_1)$ ,  $[c_1, c_2)$ , ...,  $[c_{k-1}, c_k)$  unterteilt werden. Diese Intervalle bilden jeweils eine zugehörige Klasse von  $1, \dots, k$ . Die Anzahl der Klassen sowie die Intervalllänge jeder Klasse können tatsächlich beliebig gewählt werden, solange sie das ursprüngliche Intervall vollständig überdecken und sich nicht überlappen. In der Regel werden jedoch die Längen der Intervalle gleichmäßig aufgeteilt. Die Auswahl der Intervallgröße ist entscheidend, um ein sinnvolles Histogramm zu erstellen. Zu kleine Intervalle erzeugen Rauschen, zu große verschleiern Details.

Für jede Klasse wird die absolute Häufigkeit  $h_j, j = 1, \dots, k$  (Anzahl der vorkommenden Werte innerhalb des zugehörigen Intervalls) gezählt. Dividiert man die absoluten Häufigkeiten durch den Stichprobenumfang  $n$ , so erhält man die relativen Häufigkeiten,  $f_j = h_j/n, j = 1, \dots, k$ . Auf jedem Intervall wird nun eine Säule aufgetragen. Dementsprechend stimmt die Breite jeder Säule mit der Intervalllänge  $d_j = c_j - c_{j-1}, j = 1, \dots, k$  überein. Um die Fläche der Säule proportional zu den Häufigkeitsdaten darzustellen, beträgt ihre Höhe  $h_j/d_j$  bzw.  $f_j/d_j$ .

## 3.2. Multivariate Verfahren

Zur Beschreibung der folgenden Methoden wird von zwei metrischen Variablen  $X$  und  $Y$  ausgegangen. Dabei sind ihre beobachteten Werte miteinander verbunden und als gepaarte Daten  $(x_1, y_1), \dots, (x_n, y_n)$  dargestellt.

**Korrelationskoeffizient nach Bravais-Pearson:** Der empirische Korrelationskoeffizient nach Bravais-Pearson (Pearson-Korrelationskoeffizient) ist ein Maß für die Stärke und Richtung der linearen Beziehung zwischen zwei Variablen.

Für  $n$  gepaarte Daten  $(x_1, y_1), \dots, (x_n, y_n)$  ist es definiert als:

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{s_{XY}}{s_X s_Y},$$

wobei

- $(x_i, y_i)$  die Messwerte zweier stetiger Merkmale sind und  $n$  der Stichprobenumfang ist,
- $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  der Mittelwert des Merkmals  $X$  bzw.  $Y$  ist,
- $s_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ ,  $s_Y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$  die Standardabweichung des Merkmals  $X$  bzw.  $Y$  ist und  $s_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$  die empirische Kovarianz bezeichnet.

Es wird gezeigt, dass  $-1 < r_{XY} < 1$  gilt (Fahrmeier et al., 1987, S. 136-140). Die Interpretation von Werten nahe +1 bzw. -1 deutet auf eine starke positive bzw. negative Korrelation hin, während Werte um Null auf eine schwache oder gar keine Korrelation hinweisen.

**Streudiagramm:** In einem Streudiagramm erfolgt die gemeinsame Darstellung der Messwerte  $(x_i, y_i)$  zweier stetiger Merkmale in einem  $(x - y)$ -Koordinatensystem mittels Punkte, Kreuze oder sonstige Symbole. Das Streudiagramm ermöglicht eine übersichtliche Darstellung von Zusammenhängen zwischen zwei oder mehreren Variablen, sodass sich Muster, Trends und Ausreißer identifizieren lassen. (Fahrmeier et al., 1987, S. 129-130)



## 4. Statistische Auswertung

Kapitel 4 widmet sich der Datenanalyse. Zur Synchronisierung mit der Verbrechensrate werden die Ausgaben sowie die Anzahl der kriminalistischer Beobachtungen, der Anwälte und der Staatsdiener auf eine Bevölkerung von 100.000 Einwohnern nomiert, indem der ursprüngliche Wert durch die Anzahl der Einwohner geteilt und mit 100.000 multipliziert wird. Die Bevölkerungsdichte wird auch berechnet, indem die Anzahl der Einwohner durch die Fläche geteilt wird. Abgesehen davon werden keine weiteren Datenmanipulationen durchgeführt.

### 4.1. Häufigkeitsverteilungen der Variablen

Die Kennzahlen werden in Tabelle 1 zusammengefasst. Die Anzahl der Personen unter strafrechtlicher Überwachung ist äußerst gering. Sogar der Maximalwert liegt noch unter vier. Sowohl der Median als auch der Mittelwert zeigen, dass die Dichte der Anwälte etwa 100 Personen niedriger ist als die Dichte der Beamten. Allerdings sind beide Variablen um ein Vielfaches niedriger im Vergleich zur Kriminalitätsrate. Im Allgemeinen weisen Variablen mit einem großen Lagemaß ebenfalls eine hohe Streuung auf.

| Variable                               | Min     | Median  | Mittelwert | Max     | SD      | IQR     |
|--|---------|---------|------------|---------|---------|---------|
| Nomierte Ausgaben (in \$1.000)         | 8.86    | 15.20   | 17.88      | 69.94   | 11.35   | 6.20    |
| Verbrechensrate                        | 2253.00 | 4549.00 | 4801.84    | 8339.00 | 1383.28 | 1678.50 |
| Anwalt-Dichte                          | 144.47  | 232.46  | 323.17     | 4565.76 | 608.77  | 57.93   |
| Staatsdiener-Dichte                    | 267.74  | 387.38  | 424.02     | 1274.12 | 144.62  | 93.81   |
| Kriminalistische Beobachtungsdichte    | 0.36    | 0.93    | 1.02       | 3.75    | 0.55    | 0.44    |
| Bevölkerungsdichte pro km <sup>2</sup> | 0.31    | 30.07   | 120.86     | 3514.12 | 489.58  | 50.02   |

Tabelle 1: Statistische Zusammenfassung der untersuchten Variablen

Die große Abweichung zwischen Mittelwert und Median kann auf die starke Beeinflussung durch Extremwerte hinweisen. Die Boxplots der untersuchten Variablen veranschaulichen dieses Phänomen (siehe Abbildung 1). Bei allen Variablen treten Ausreißer lediglich rechtsseitig und sehr weit entfernt zu den übrigen Punkten auf. Insbesondere bei der Bevölkerungs- und Anwaltsdichte wird die Box so dünn, dass ihre Bestandteile nicht mehr zu unterscheiden sind. Es fällt auf, dass sämtliche Ausreißer mit Maximalwerten aus DC stammen. Die Histogramme in Abbildung 3 zeigen, dass alle Verteilungen unimodal und rechtsschief sind. Zur einfacheren Visualisierung werden die Daten aus DC dabei ausgelassen. Die Verteilung der Anwaltsdichte wird dadurch flacher, während die Verteilung der Bevölkerungsdichte spitz bleibt.

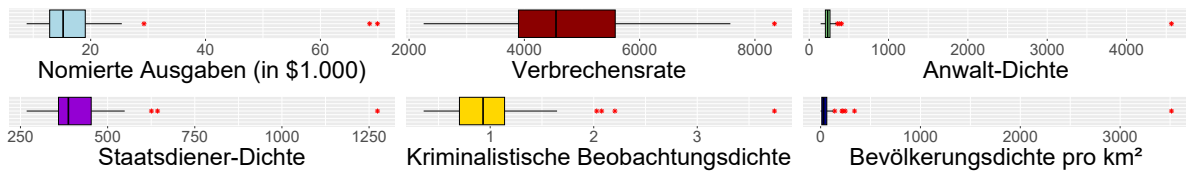


Abbildung 1: Boxplots der untersuchten Variablen

## 4.2. Zusammenhänge zwischen den Variablen

Im Folgenden werden die bivariaten Zusammenhänge zwischen den oben betrachteten Variablen untersucht. Dazu wird für jedes Variablenpaar ein Streudiagramm erstellt und der Korrelationskoeffizient nach Pearson berechnet (siehe Abbildung 2). Zusätzlich zu den Daten aus DC werden auch die Daten aus AK in diesem Analyseprozess nicht berücksichtigt. Wenn man diese beiden Datenpunkte in den Streudiagrammen darstellt, kann man leicht erkennen, dass sie sich von der Wolke der restlichen Punkte abheben (vergleiche mit den Streudiagrammen in Abbildung 4). Auch die Korrelationskoeffizienten ändern sich stark. Insbesondere steigen die Korrelationskoeffizienten zwischen der Bevölkerungsdichte und den anderen Variablen deutlich an. Dies ist darauf zurückzuführen, dass dieses Korrelationsmaß empfindlich auf Extremwerte reagiert. Und wie bereits erwähnt, dominiert bei der Bevölkerungsdichte der Extremwert die Verteilung.



Abbildung 2: Streudiagramme und Korrelationskoeffizienten der untersuchten Variablen

Wie in Abbildung 2 dargestellt, sind die Werte der Korrelationskoeffizienten alle größer als Null. Das heißt, die Korrelationen zwischen den Variablen sind positiv. Am größten sind die Korrelationen zwischen Staatsausgaben und der Beamtendichte, der Kriminalitätsrate und der Anwaltsdichte. Dabei liegt die erste nahe am Schwellenwert der starken Korrelation. Die folgenden zwei gelten ebenfalls als mittlere. Bei Betrachtung der entsprechenden Streudiagramme fällt auch ein visueller Eindruck auf, dass die Datenpunkte von der linken unteren Ecke bis zur rechten oberen Ecke des Diagramms verstreut sind.

## 5. Zusammenfassung

In Anbetracht komplexer sozialer Situationen und limitierter staatlicher Ressourcen transformiert sich die Bereitstellung finanzieller Mittel für das Rechtssystem zu einer politischen Herausforderung. Der Fokus des Projekts liegt auf der Identifizierung von Faktoren, die die Ausgaben der US-Bundesstaaten für die Kriminalitätsbekämpfung beeinflussen. Der verwendete Datensatz enthielt Informationen zu Ausgaben, Kriminalitätsraten sowie zur Anzahl der Anwälte bzw. öffentlichen Angestellten aus allen Bundesstaaten der USA, einschließlich District of Columbia. Darüber hinaus wurden auch Angaben zur Bevölkerung und Fläche bereitgestellt.

Bei der Analyse der unabhängigen Daten wurde festgestellt, dass die Verteilungen der Variablen alle eine Rechtsschiefheit aufweisen. Des Weiteren wurde der Auftritt von Ausreißer aufgezeichnet. Es war bemerkenswert, dass sämtliche Extremwerte dem DC angehören, der im Wesentlichen eine Sonderverwaltungszone innerhalb der USA ist. Durch die Untersuchung der Korrelationen zwischen Variablen konnten potenzielle Faktoren identifiziert werden, die die Staatsausgaben beeinflussen. Hierbei sind die Kriminalitätsrate, die Anwaltsdichte sowie die Beamtendichte zu berücksichtigen.

Der vorliegende Datensatz hat Auffälligkeiten bei der kriminelle Überwachung aufgewiesen. Fehlende Zeitangaben erschwerten die Interpretation der Variablen. Auch die Korrelationsanalyse ist nicht in der Lage, die Ausreißer zu erklären. Allerdings können die Ergebnisse dieses Projekts potenziell den Ausgangspunkt für weitere kausale Studien bilden. Darüber hinaus besteht die Möglichkeit, die Aussagekraft der Forschung durch den Erwerb von Fachwissen zu erhöhen.

## Literatur

L. Fahrmeier, C. Heumann, R. Künstler, I. Pigeot, and G. Tutz. *Statistik: Der Weg zur Datenanalyse*. Springer-Verlag, New York, 1987. (8. Auflage, 2016),.

ggplot2. A box and whiskers plot (in the style of tukey), 2024. URL [https://ggplot2.tidyverse.org/reference/geom\\_boxplot.html](https://ggplot2.tidyverse.org/reference/geom_boxplot.html). besucht am 20.04.2025.

R Core Team. R: A language and environment for statistical computing, 2024. URL <https://www.R-project.org/>.

Hadley Wickham, Danielle Navarro, and Thomas Lin Pedersen. *ggplot2: Elegant Graphics for Data Analysis (3e)*. Springer-Verlag, work-in-progress 3rd edition, 2016. URL <https://ggplot2-book.org/>.

## A. Zusätzliche Grafiken

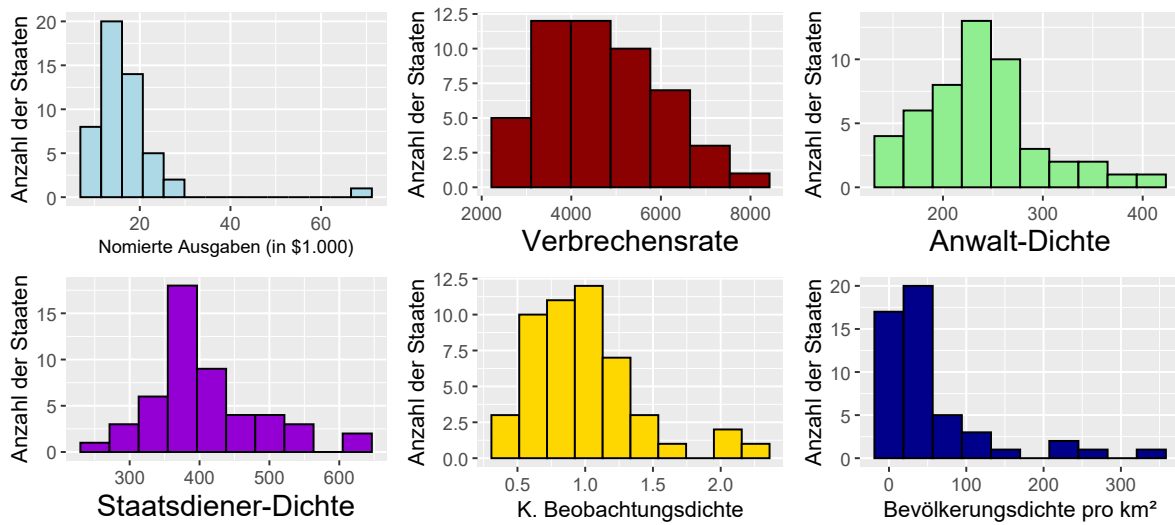


Abbildung 3: Histogramme der untersuchten Variablen (ohne DC)

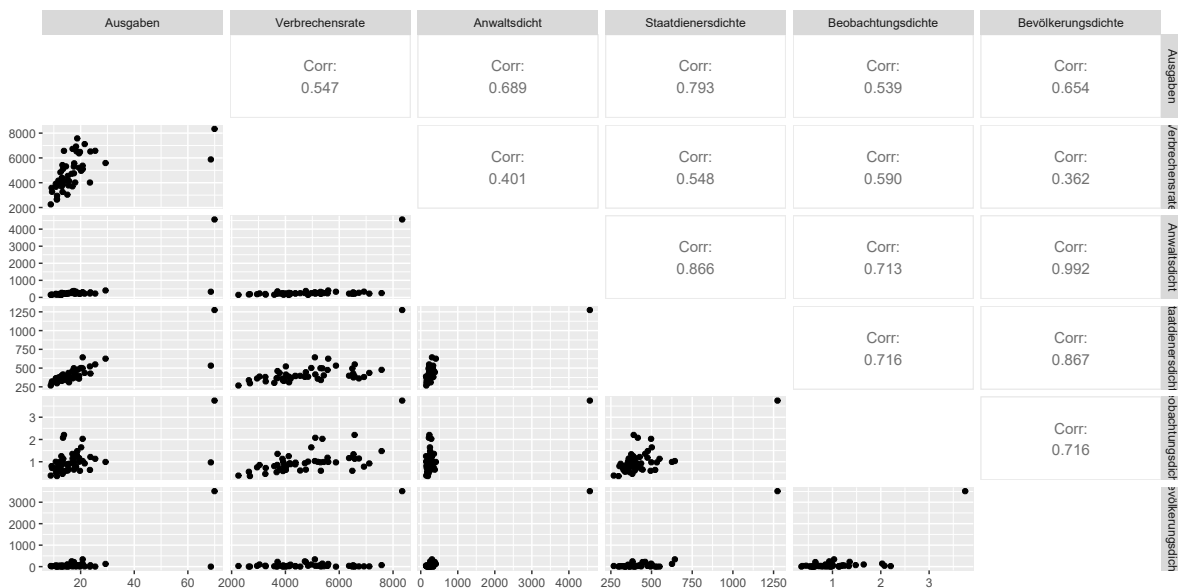


Abbildung 4: Streudiagramme und Korrelationskoeffizienten der untersuchten Variablen