

Corpus der Entscheidungen
des
Bundesgerichtshofs
(CE-BGH)

CODEBOOK

Version 2024-09-25



DOI: [10.5281/zenodo.12814022](https://doi.org/10.5281/zenodo.12814022)

Titel	Corpus der Entscheidungen des Bundesgerichtshofs
Abkürzung	CE-BGH
Autor	Seán Fobbe
Version	2024-09-25
Download	https://doi.org/10.5281/zenodo.12814022
Lizenz	CC0 1.0 Universal

Zitiervorschlag

Seán Fobbe (2024). Corpus der Entscheidungen des Bundesgerichtshofs (CE-BGH). Version 2024-09-25. Zenodo. DOI: [10.5281/zenodo.12814022](https://doi.org/10.5281/zenodo.12814022).

Digital Object Identifier (DOI): Concept DOI und Version DOI

Soweit nicht anders angegeben ist die DOI immer eine »Version DOI« und bezieht sich nur auf eine bestimmte Version des Datensatzes. Sie verweist daher nur auf Version 2024-09-25. Für das Gesamtkonzept dieses Datensatzes steht eine »Concept DOI« zur Verfügung, die auf der Zenodo-Seite jeder Version unter »Cite all versions?« zu finden ist. Sie lautet [10.5281/zenodo.3942742](https://doi.org/10.5281/zenodo.3942742). Die »Concept DOI« verlinkt immer die aktuellste Version.

Urheberrecht

Der Datensatz und dieses Dokument sind unter einer **Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication Lizenz** veröffentlicht. Ich stelle den Datensatz und das Codebook vollständig gemeinfrei und verzichte weltweit auf alle damit verbundenen Urheberrechte, einschließlich aller ähnlichen Rechte, soweit dies gesetzlich möglich ist.

Sie können die Werke kopieren, modifizieren, verteilen und aufführen ohne um Erlaubnis bitten zu müssen, selbst für kommerzielle Zwecke. Patente und Markenschutzrechte bleiben von CC0 unberührt. CC0 hat auch keine Auswirkungen auf etwaige Datenschutz- oder Persönlichkeitsrechte. Jegliche Haftung für die Benutzung dieses Werkes ist ausgeschlossen, bis zu dem maximalen Umfang in dem dies gesetzlich möglich ist.

Wenn Sie diese Werke nutzen oder zitieren sollten Sie nicht den Eindruck erwecken, der Autor unterstütze ihre Nutzung.

Dies ist nur eine unverbindliche deutsche Zusammenfassung der Lizenz, den vollständigen und rechtsverbindlichen Lizenztext finden Sie hier: <https://creativecommons.org/publicdomain/zero/1.0/legalcode>

Disclaimer

Dieser Datensatz ist eine private wissenschaftliche Initiative und steht in keiner Verbindung zu Behörden, Gerichten oder anderen amtlichen Stellen der Bundesrepublik Deutschland.

Inhaltsverzeichnis

1	Einführung	4
2	Nutzung	5
2.1	CSV-Dateien	5
2.2	TXT-Dateien	5
3	Konstruktion	6
3.1	Beschreibung des Datensatzes	6
3.2	Datenquellen	6
3.3	Sammlung der Daten	6
3.4	Source Code und Compilation Report	6
3.5	Grenzen des Datensatzes	8
3.6	Urheberrechtsfreiheit von Rohdaten und Datensatz	8
3.7	Metadaten	8
3.7.1	Allgemein	8
3.7.2	Schema für die Dateinamen	9
3.7.3	Beispiel eines Dateinamens	9
3.8	Qualitätsprüfung	9
3.9	Grafische Darstellung	9
4	Varianten und Zielgruppen	10
5	Variablen	12
5.1	Datenstruktur	12
5.2	Hinweise	13
5.3	Erläuterung der Variablen	13
6	Registerzeichen	20
7	Präsident:innen	24
7.1	Hinweise	24
7.2	Lebensdaten	24
7.3	Dienstalter und Lebensalter	24
8	Vize-Präsident:innen	25
8.1	Hinweise	25
8.2	Lebensdaten	25
8.3	Dienstalter und Lebensalter	26
9	Linguistische Kennzahlen	27
9.1	Erläuterung der Kennzahlen und Diagramme	27
9.2	Werte der Kennzahlen	27
9.3	Verteilung Zeichen	28
9.4	Verteilung Tokens	28
9.5	Verteilung Typen	29
9.6	Verteilung Sätze	29
10	Zitationsnetzwerk des Bundesgerichtshofs (Beta)	30
10.1	Überblick	30

10.2 Technische Hinweise	30
10.3 Methodik Aktenzeichen	31
10.4 Methodik BGHZ und BGHSt	31
10.4.1 Erste Stufe	31
10.4.2 Zweite Stufe	32
10.4.3 Grenzen	33
10.5 Visualisierungen des Zitationsnetzwerks	33
11 Inhalt des Korpus	35
11.1 Zusammenfassung	35
11.2 Nach Typ der Entscheidung	35
11.3 Nach Spruchkörper (Aktenzeichen)	36
11.4 Nach Spruchkörper (Datenbank)	38
11.5 Nach Registerzeichen	40
11.6 Nach Präsident:in	43
11.7 Nach Vize-Präsident:in	44
11.8 Nach Entscheidungsjahr	45
11.9 Nach Eingangsjahr (ISO)	47
12 Dateigrößen	49
13 Kryptographische Signaturen	50
13.1 Zwei-Phasen-Signatur	50
13.2 Persönliche GPG-Signatur	50
14 Changelog	51
14.1 Version 2024-09-25	51
14.2 Version 2023-03-10	51
14.3 Version 2022-08-16	51
14.4 Version 2022-02-12	52
14.5 Version 2021-04-27	52
14.6 Version 2020-07-09	53
15 Parameter für strenge Replikationen	54
Literaturverzeichnis	56

1 Einführung

Der **Bundesgerichtshof (BGH)** ist einer der fünf obersten Gerichtshöfe des Bundes und steht an der Spitze der ordentlichen Gerichtsbarkeit der Bundesrepublik Deutschland (Art. 95 Abs. 1 GG, §§ 12, 13 GVG).¹ Der BGH ist die höchste Instanz in Zivil- und Strafsachen, sowie in einigen ihm zugewiesenen Spezialgebieten. Er wurde am 1. Oktober 1950 errichtet und hat seinen Sitz in Karlsruhe (§ 123 GVG). Der 5. und 6. Strafsenat sind allerdings in Leipzig beheimatet.

Im Jahr 2022 am BGH eingerichtet sind 13 Zivilsenate, 6 Strafsenate, und 8 Spezialsenate (6 berufsrechtliche Senate, Kartellsenat und Landwirtschaftsenat), sowie ein Großer Zivilsenat, ein Großer Strafsenat und die Vereinigten Großen Senate.² Ein Senat hat grundsätzlich 7 bis 9 Mitglieder, entscheidet aber als Spruchkörper in Senatsgruppen von 5 Mitgliedern einschließlich des/der Vorsitzenden (§ 139 Abs. 1 GVG).

Die überwiegende Anzahl der Verfahren vor dem BGH sind Revisionen, d.h. die Überprüfung von Entscheidungen unterer Instanzen (Landgerichte oder Oberlandesgerichte, Amtsgericht nur bei Sprungrevision) auf Rechtsfehler ohne erneute Beweisaufnahme. In Zivilsachen ist der BGH zuständig für Revisionen, Sprungrevisionen, Rechtsbeschwerden und Sprungrechtsbeschwerden (§ 133 GVG). In Strafsachen ist er zuständig für Revisionen, Beschwerden gegen Beschlüsse und Verfügungen der Oberlandesgerichte, Beschwerden gegen Verfügungen des Ermittlungsrichters am BGH und Rügen der Besetzung eines Oberlandesgerichts (§ 135 GVG).

Wieso dieser Datensatz? Die quantitative Analyse von juristischen Texten, insbesondere denen des BGH, ist in den deutschen Rechtswissenschaften ein noch junges und kaum bearbeitetes Feld.³ Zu einem nicht unerheblichen Teil liegt dies auch daran, dass die Anzahl an frei nutzbaren Datensätzen außerordentlich gering ist.

Die meisten hochwertigen Datensätze lagern (fast) unerreichbar in kommerziellen Datenbanken und sind wissenschaftlich gar nicht oder nur gegen Entgelt zu nutzen. Frei verfügbare Datenbanken wie *Opinio Iuris*⁴ und *openJur*⁵ verbieten ausdrücklich das maschinelle Auslesen der Rohdaten. Wissenschaftliche Initiativen wie der Juristische Referenzkorpus (JuReKo) sind nach jahrelanger Arbeit hinter verschlossenen Türen verschwunden.

In einem funktionierenden Rechtsstaat muss die Rechtsprechung öffentlich, transparent und nachvollziehbar sein. Im 21. Jahrhundert bedeutet dies auch, dass sie systematischer Überprüfung mittels quantitativen Analysen zugänglich sein muss. Der Erstellung und Aufbereitung des Datensatzes liegen daher die Prinzipien der allgemeinen Verfügbarkeit durch Urheberrechtsfreiheit, strenge Transparenz und vollständige wissenschaftliche Reproduzierbarkeit zugrunde. Die FAIR-Prinzipien (Findable, Accessible, Interoperable and Reusable) für freie wissenschaftliche Daten inspirieren sowohl die Konstruktion, als auch die Art der Publikation.⁶

¹ Die »ordentliche Gerichtsbarkeit« ist eine historische gewachsene Bezeichnung. Früher war die Verwaltungsgerichtsbarkeit nicht mit unabhängigen Richtern, sondern mit Verwaltungsbeamten besetzt und daher »außerordentlich«. Die mit unabhängigen Richtern besetzten Gerichte wurden als »ordentlich« bezeichnet.

² Geschäftsverteilungsplan des Bundesgerichtshofs für das Jahr 2021.

³ Besonders positive Ausnahmen finden sich unter: <https://www.quantitative-rechtswissenschaft.de/>
⁴ <https://opinioiuris.de/>

⁵ <https://openjur.de/>

⁶ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

2 Nutzung

Die Daten sind in offenen, interoperablen und weit verbreiteten Formaten (CSV, TXT, PDF) veröffentlicht. Sie lassen sich grundsätzlich mit allen modernen Programmiersprachen (z.B. Python oder R), sowie mit grafischen Programmen nutzen.

Wichtig: Nicht vorhandene Werte sind sowohl in den Dateinamen als auch in der CSV-Datei mit "NA" codiert.

2.1 CSV-Dateien

Am einfachsten ist es die **CSV-Dateien** einzulesen. CSV⁷ ist ein einfaches und maschinell gut lesbares Tabellen-Format. In diesem Datensatz sind die Werte komma-separiert. Jede Spalte entspricht einer Variable, jede Zeile einer Entscheidung. Die Variablen sind unter Punkt 5 genauer erläutert.

Zum Einlesen empfehle ich für **R** das package **data.table** (via CRAN verfügbar). Dessen Funktion **fread()** ist etwa zehnmal so schnell wie die normale **read.csv()**-Funktion in Base-R. Sie erkennt auch den Datentyp von Variablen sicherer. Ein Beispiel:

```
library(data.table)
dt.bgh <- fread("filename.csv")
```

2.2 TXT-Dateien

Die **TXT-Dateien** inklusive Metadaten können zum Beispiel mit **R** und dem package **readtext** (via CRAN verfügbar) eingelesen werden. Ein Vorschlag:

```
library(readtext)
df.bgh <- readtext("./*.txt",
  docvarsfrom = "filenames",
  docvarnames = c("gericht",
                  "spruchkoerper_db",
                  "leitsatz",
                  "datum",
                  "spruchkoerper_az",
                  "registerzeichen",
                  "eingangsnummer",
                  "eingangsjahr_az",
                  "zusatz_az",
                  "name",
                  "kollision"),
  dvsep = "_",
  encoding = "UTF-8")
```

⁷ Das CSV-Format ist in RFC 4180 definiert, siehe <https://tools.ietf.org/html/rfc4180>

3 Konstruktion

3.1 Beschreibung des Datensatzes

Dieser Datensatz ist eine digitale Zusammenstellung von möglichst allen begründeten Entscheidungen, die auf der amtlichen Internetpräsenz des Bundesgerichtshofs (BGH) am jeweiligen Stichtag veröffentlicht waren. Die Stichtage für jede Version entsprechen exakt der Versionsnummer.

Zusätzlich zu den aufbereiteten maschinenlesbaren Formaten (TXT und CSV) sind die PDF-Rohdaten enthalten, damit Analyst:innen gegebenenfalls ihre eigene Konvertierung vornehmen können. Die PDF-Rohdaten wurden inhaltlich nicht verändert und nur die Dateinamen angepasst, um die Lesbarkeit für Mensch und Maschine zu verbessern.

Speziell an Praktiker:innen richten sich die PDF-Sammlungen aller Leitsatzentscheidungen und aller mit Namen versehenen Entscheidungen.

3.2 Datenquellen

Datenquelle	Fundstelle
Primäre Datenquelle	https://www.bundesgerichtshof.de
Source Code	https://doi.org/10.5281/zenodo.12814023
Personendaten	https://doi.org/10.5281/zenodo.4568682
Registerzeichen	https://doi.org/10.5281/zenodo.4569564

Die Personendaten stammen aus folgendem Datensatz: »Seán Fobbe and Tilko Swalve (2021). Presidents and Vice-Presidents of the Federal Courts of Germany (PVP-FCG). Version 2021-04-08. Zenodo. DOI: 10.5281/zenodo.4568682«.

Die Tabelle der Registerzeichen und der ihnen zugeordneten Verfahrensarten stammt aus dem folgenden Datensatz: “Seán Fobbe (2021). Aktenzeichen der Bundesrepublik Deutschland (AZ-BRD). Version 1.0.1. Zenodo. DOI: 10.5281/zenodo.4569564.”

3.3 Sammlung der Daten

Die Daten wurden unter Beachtung des Robot Exclusion Standard (RES) gesammelt. Der Abruf geschieht ausschließlich über TLS-verschlüsselte Verbindungen. Die Entscheidungen sind laut dem Gericht anonymisiert, aber ungekürzt.

3.4 Source Code und Compilation Report

Der gesamte Source Code — sowohl für die Erstellung des Datensatzes, als auch für dieses Codebook — ist öffentlich einsehbar und dauerhaft erreichbar im wissenschaftlichen Archiv des CERN unter dieser Adresse hinterlegt: <https://doi.org/10.5281/zenodo.12814023>

Mit jeder Kompilierung des vollständigen Datensatzes wird auch ein umfangreicher **Compilation Report** in einem attraktiv designten PDF-Format erstellt (ähnlich diesem

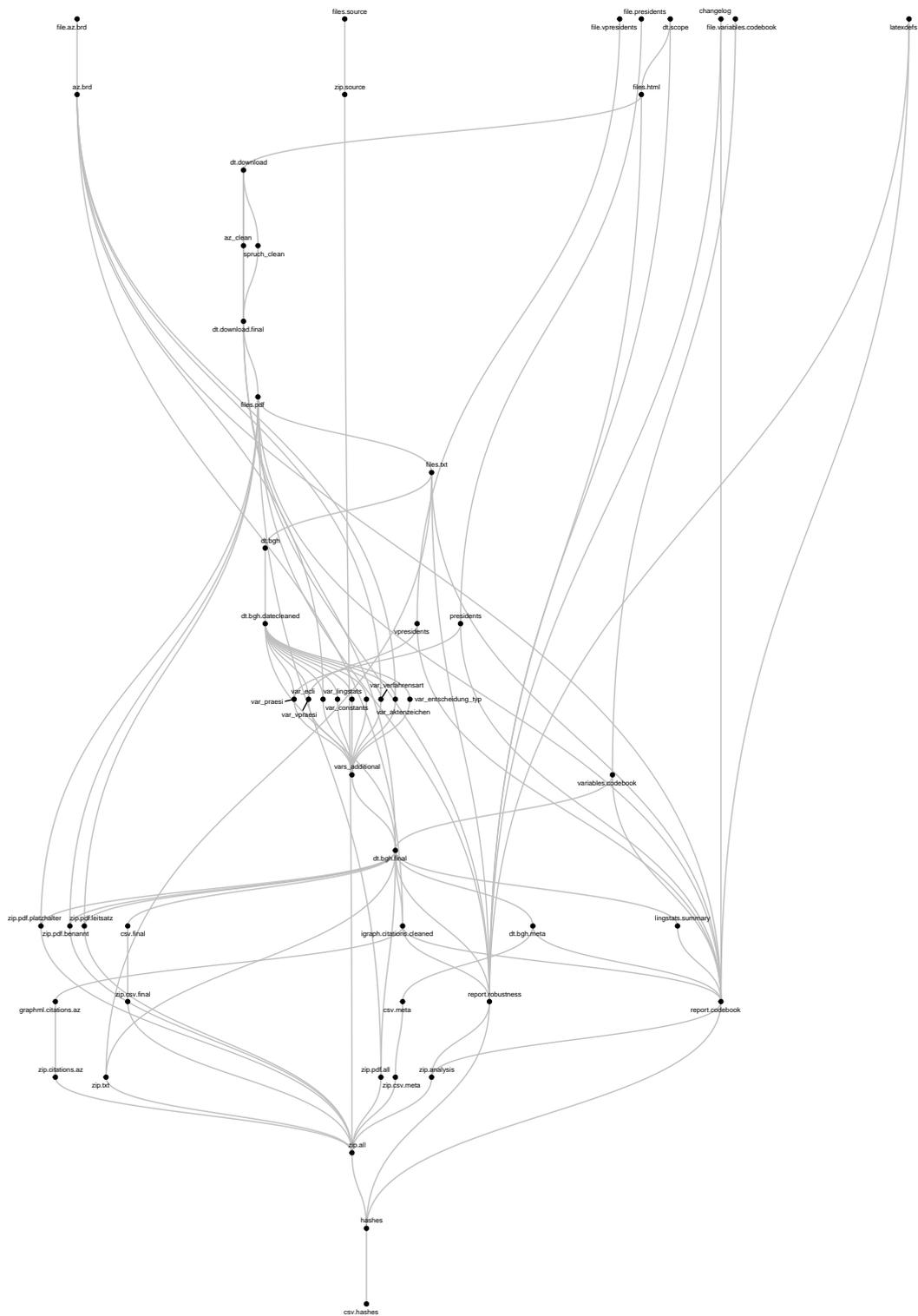


Abbildung 1: Der vollständige Prozess der Datensatz-Kompilierung.

Codebook). Der Compilation Report enthält den kommentierten Source Code für die Daten-Pipeline, dokumentiert relevante Rechenergebnisse, gibt sekundengenaue Zeitstempel an und ist mit einem klickbaren Inhaltsverzeichnis versehen. Er ist zusammen mit dem Source Code hinterlegt. Wenn Sie sich für Details der Herstellung interessieren, lesen Sie diesen bitte zuerst.

3.5 Grenzen des Datensatzes

Nutzer:innen sollten folgende wichtige Grenzen beachten:

- Der Datensatz enthält nur das, was das Gericht auch tatsächlich veröffentlicht, nämlich begründete Entscheidungen (*publication bias*). Platzhalter für Nichtannahme- und Verwerfungsbeschlüsse ohne Begründung wurden automatisiert entfernt, sofern vorhanden.
- Es kann aufgrund technischer Grenzen bzw. Fehler sein, dass manche — im Grunde verfügbare — Entscheidungen nicht oder nicht korrekt abgerufen werden (*automation bias*).
- Es werden nur PDF-Dateien abgerufen (*file type bias*). Manche Entscheidungen sind nur als HTML verfügbar.
- Erst ab dem 1. Januar 2000 sind begründete Entscheidungen des Bundesgerichtshofs einigermaßen vollständig veröffentlicht (*temporal bias*). Die Frequenztabellen geben hierzu genauer Auskunft.

3.6 Urheberrechtsfreiheit von Rohdaten und Datensatz

An den Entscheidungstexten und amtlichen Leitsätzen besteht gem. § 5 Abs. 1 UrhG kein Urheberrecht, da sie amtliche Werke sind. § 5 UrhG ist auf amtliche Datenbanken analog anzuwenden (BGH, Beschluss vom 28.09.2006, I ZR 261/03, »Sächsischer Ausschreibungsdienst«).

Alle eigenen Beiträge (z.B. durch Zusammenstellung und Anpassung der Metadaten) und damit den gesamten Datensatz stelle ich gemäß einer *CC0 1.0 Universal Public Domain Lizenz* vollständig urheberrechtsfrei.

3.7 Metadaten

3.7.1 Allgemein

Die Metadaten in den Dateinamen sind größtenteils unverändert von den jeweiligen Datenbankeinträgen aus der amtlichen Datenbank des Bundesgerichtshofs entnommen. Berechnet und hinzugefügt wurden durch den Autor des Datensatzes eine Reihe weitere Variablen, sowie in den Dateinamen der PDF/TXT-Dateien Unter- und Trennstriche, um die Maschinenlesbarkeit zu erleichtern. Der volle Satz an Metadaten ist nur in den CSV-Dateien enthalten. Alle hinzugefügten Metadaten sind vollständig maschinenlesbar dokumentiert. Sie sind entweder im Source Code enthalten, mit dem Source Code zusammen dokumentiert oder über dauerhaft stabile Identifikatoren (z.B. DOI) zitiert.

Die Dateinamen der PDF- und TXT-Dateien enthalten Gerichtsname, Spruchkörperangabe aus der Datenbank, die Angabe ob es sich um eine Leitsatzentscheidung handelt, Datum, das offizielle Aktenzeichen, einen Zusatz zum Aktenzeichen, ggf. einen besonderen Namen und eine durch den Autor des Datensatzes generierte Kollisions-ID.

3.7.2 Schema für die Dateinamen

[gericht]_[spruchkoerper_db]_[leitsatz]_[datum]_[spruchkoerper_az]_
[registerzeichen]_[eingangsnummer]_[eingangsjahr_az]_[zusatz_az]_
[name]_[kollision]

3.7.3 Beispiel eines Dateinamens

BGH_Zivilsenat-1_LE_2007-10-04_I_ZR_22_05_NA_Umsatzsteuerhinweis_0.pdf

3.8 Qualitätsprüfung

Die Typen der Variablen wurden mit *regular expressions* strikt validiert. Die möglichen Werte der jeweiligen Variablen wurden zudem durch Frequenztabellen und Visualisierungen auf ihre Plausibilität geprüft. Insgesamt werden zusammen mit jeder Kompilierung Dutzende Tests zur Qualitätsprüfung durchgeführt. Alle Ergebnisse der Qualitätsprüfungen sind aggregiert im Compilation Report und einzeln im Archiv »ANALYSE« zusammen mit dem Datensatz veröffentlicht.

3.9 Grafische Darstellung

Die Robenfarbe der Richter:innen des Bundesgerichtshofs ist »karmesinrot«. Der Hex-Wert hierfür ist vermutlich #7e0731. Das ist besonders bei der Erstellung thematisch passender Diagrammen hilfreich. Alle im Compilation Report und diesem Codebook präsentierten Diagramme sind in diesem karmesinrot gehalten.

4 Varianten und Zielgruppen

Dieser Datensatz ist in verschiedenen Varianten verfügbar, die sich an unterschiedliche Zielgruppen richten. Zielgruppe sind nicht nur quantitativ forschende Rechtswissenschaftler:innen, sondern auch traditionell arbeitende Jurist:innen. Idealerweise müssen quantitative Methoden ohnehin immer durch qualitative Interpretation, Theoriebildung und kritische Auseinandersetzung verstärkt werden (*mixed methods approach*).

Lehrende werden von den vorbereiteten Tabellen und Diagrammen besonders profitieren, die bei der Erläuterung der Charakteristika der Daten hilfreich sein können und Zeit im universitären Alltag sparen. Alle Tabellen und Diagramme liegen auch als separate Dateien vor um sie einfach z.B. in Präsentations-Folien oder Handreichungen zu integrieren.

Variante	Zielgruppe und Beschreibung
PDF	Traditionelle juristische Forschung. Die PDF-Dokumente wie sie vom BGH auf der amtlichen Webseite bereitgestellt werden, jedoch verbessert durch semantisch hochwertige Dateinamen, die der leichteren Auffindbarkeit von Entscheidungen dienen. Die Dateinamen sind so konzipiert, dass sie auch für die traditionelle qualitative juristische Arbeit einen erheblichen Mehrwert bieten. Im Vergleich zu den CSV-Dateien enthalten die Dateinamen nur einen reduzierten Umfang an Metadaten, um Kompatibilitätsprobleme zu vermeiden und die Lesbarkeit zu verbessern. Neben dem vollen Datensatz sind für Praktiker:innen auch Varianten aufbereitet, die nur <i>Leitsatzentscheidungen</i> oder nur <i>besonders wichtige Entscheidungen mit Namen</i> enthalten.
CSV_Datensatz	Legal Tech/Quantitative Forschung. Diese CSV-Datei ist die für statistische Analysen empfohlene Variante des Datensatzes. Sie enthält den Volltext aller Entscheidungen, sowie alle in diesem Codebook beschriebenen Metadaten. Jede Spalte entspricht einer Variable, jede Zeile einer Entscheidung. Über Zeilenumbrüche getrennte Wörter wurden zusammengefügt.
CSV_Metadaten	Legal Tech/Quantitative Forschung. Wie die vorige CSV-Variante, nur ohne die Entscheidungstexte. Sinnvoll für Analyst:innen, die sich nur für die Metadaten interessieren und Speicherplatz sparen wollen. Jede Spalte entspricht einer Variable, jede Zeile einer Entscheidung.
TXT	Subsidiär für alle Zielgruppen. Diese Variante enthält die vollständigen aus den PDF-Dateien extrahierten Entscheidungstexte, aber nur einen reduzierten Umfang an Metadaten, der dem der PDF-Dateien entspricht. Die TXT-Dateien sind optisch an das Layout der PDF-Dateien angelehnt. Geeignet für qualitativ arbeitende Forscher:innen, die nur wenig Speicherplatz oder eine langsame Internetverbindung zur Verfügung haben oder für quantitativ arbeitende Forscher:innen, die beim Einlesen der CSV-Dateien Probleme haben. Über Zeilenumbrüche getrennte Wörter wurden <i>nicht</i> zusammengefügt.

Variante	Zielgruppe und Beschreibung
Netzwerke	Alle Lehrenden und Forschenden. EXPERIMENTELL! Das Zitationsnetzwerk des BGH. Umfasst sind nur Zitate von Aktenzeichen des BGH zu Aktenzeichen des BGH (d.h. zwischen Verfahren). Eine Auflösung auf Ebene einzelner Entscheidungen ist geplant. Kann mit statistischer Software oder aber auch GUI-Anwendungen wie Gephi analysiert werden.
ANALYSE	Alle Lehrenden und Forschenden. Dieses Archiv enthält alle während dem Kompilierungs- und Prüfprozess erstellten Tabellen (CSV) und Diagramme (PDF, PNG) im Original. Sie sind inhaltsgleich mit den in diesem Codebook verwendeten Tabellen und Diagrammen. Das PDF-Format eignet sich besonders für die Verwendung in gedruckten Publikationen, das PNG-Format besonders für die Darstellung im Internet. Analyst:innen mit fortgeschrittenen Kenntnissen in R können auch auf den Source Code zurückgreifen. Empfohlen für Nutzer:innen die einzelne Inhalte aus dem Codebook für andere Zwecke (z.B. Präsentationen, eigene Publikationen) weiterverwenden möchten.

5 Variablen

5.1 Datenstruktur

```
## Classes 'data.table' and 'data.frame': 77892 obs. of 35 variables:
## $ doc_id : chr "BGH_Anwaltsenat_LE_2000-02-14_NA_AnwZB_12_99_NA_NA
_0.txt" "BGH_Anwaltsenat_LE_2000-02-14_NA_AnwZB_9_99_NA_NA_0.txt" "BGH_
Anwaltsenat_LE_2000-03-13_NA_AnwZB_28_99_NA_NA_0.txt" "BGH_Anwaltsenat_LE
_2000-05-29_NA_AnwZB_33_99_NA_NA_0.txt" ...
## $ url : chr "https://juris.bundesgerichtshof.de/cgi-bin/
rechtsprechung/document.py?Gericht=bgh&Art=en&Datum=2000&Sort=1&Seit"| __
truncated__ "https://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/
document.py?Gericht=bgh&Art=en&Datum=2000&Sort=1&Seit"| __truncated__ "https
://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&
Art=en&Datum=2000&Sort=1&Seit"| __truncated__ "https://juris.
bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&
Datum=2000&Sort=1&Seit"| __truncated__ ...
## $ gericht : chr "BGH" "BGH" "BGH" "BGH" ...
## $ datum : IDate, format: "2000-02-14" "2000-02-14" ...
## $ entscheidungsjahr: int 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000
...
## $ entscheidung_typ : chr "B" "B" "B" "B" ...
## $ leitsatz : chr "LE" "LE" "LE" "LE" ...
## $ spruchkoerper_db : chr "Anwaltsenat" "Anwaltsenat" "Anwaltsenat" "
Anwaltsenat" ...
## $ spruchkoerper_az : chr NA NA NA NA ...
## $ registerzeichen : chr "AnwZB" "AnwZB" "AnwZB" "AnwZB" ...
## $ verfahrensart : chr "Beschwerden gegen Entscheidungen eines
Anwaltsgerichtshofes" ...
## $ eingangsnummer : int 12 9 28 33 59 44 55 63 65 71 ...
## $ eingangsjahr_az : int 99 99 99 99 99 99 99 99 99 99 ...
## $ eingangsjahr_iso : num 1999 1999 1999 1999 1999 ...
## $ zusatz_az : chr NA NA NA NA ...
## $ name : chr NA NA NA NA ...
## $ kollision : int 0 0 0 0 0 0 0 0 0 0 ...
## $ aktenzeichen : chr "AnwZ(B) 12/99" "AnwZ(B) 9/99" "AnwZ(B) 28/99" "
AnwZ(B) 33/99" ...
## $ ecli : chr "ECLI:DE:BGH:2000:140200ANWZ.B.12.99.0" "ECLI:DE:
BGH:2000:140200ANWZ.B.9.99.0" "ECLI:DE:BGH:2000:130300ANWZ.B.28.99.0" "ECLI:
DE:BGH:2000:290500ANWZ.B.33.99.0" ...
## $ praesi : chr "Geiss" "Geiss" "Geiss" "Geiss" ...
## $ v_praesi : chr "Jähnke" "Jähnke" "Jähnke" "Jähnke" ...
## $ bemerkung : chr "Leitsatzentscheidung" "Leitsatzentscheidung" "
Leitsatzentscheidung" "Leitsatzentscheidung" ...
## $ bghst : num 0 0 0 0 0 0 0 0 0 0 ...
## $ bghz : num 0 0 0 0 0 0 0 0 0 0 ...
## $ bghr : num 0 0 1 0 1 0 0 0 0 0 ...
## $ nachschlagewerk : num 1 1 1 1 1 1 1 1 1 1 ...
## $ berichtigung : logi FALSE FALSE FALSE FALSE FALSE ...
## $ zeichen : int 12566 13567 10306 12738 13120 10229 18019 9681
16189 11739 ...
```

```

## $ tokens      : num  1997 2231 1602 1887 1898 ...
## $ typen      : num  646 696 556 619 705 546 827 518 807 583 ...
## $ saetze     : num  107 161 95 86 89 102 141 85 165 94 ...
## $ version    : chr   "2024-09-25" "2024-09-25" "2024-09-25" "2024-09-25"
...
## $ doi_concept : chr   "10.5281/zenodo.3942742" "10.5281/zenodo.3942742"
"10.5281/zenodo.3942742" "10.5281/zenodo.3942742" ...
## $ doi_version : chr   "10.5281/zenodo.12814022" "10.5281/zenodo.12814022"
"10.5281/zenodo.12814022" "10.5281/zenodo.12814022" ...
## $ lizenz     : chr   "Creative Commons Zero 1.0 Universal" "Creative
Commons Zero 1.0 Universal" "Creative Commons Zero 1.0 Universal" "Creative
Commons Zero 1.0 Universal" ...
## - attr(*, ".internal.selfref")=<externalptr>
## - attr(*, "sorted")= chr "doc_id"

```

5.2 Hinweise

- **Doppelte Codierung der Spruchkörper** — Für viele Urteile sind die Spruchkörper doppelt enthalten, einmal aus der Datenbank in menschenlesbarer Form (Variable »spruchkoerper_db«), einmal durch das Aktenzeichen (Variable »spruchkoerper_az«). Nicht wenige Aktenzeichen enthalten keine separate und einheitliche Angabe des Spruchkörpers, weil sich diese schon aus dem Registerzeichen ergibt, z.B. für Aktenkontrollen in Haftprüfungsverfahren oder Kartellsachen. Die Variable »spruchkoerper_db« ermöglicht hier eine einfache Selektion bestimmter Spruchkörper anhand einer einheitlichen Variable, ohne im Detail erst jedes passende Registerzeichen suchen zu müssen.
- **Abweichende Codierung der Registerzeichen** — Die Registerzeichen wurden gekürzt und Sonderzeichen entfernt um die Arbeit mit ihnen zu vereinfachen. Beachten Sie bitte die Gegenüberstellungstabelle unter Punkt 6.
- **Fehlende Werte** sind immer mit »NA« codiert.
- **Strings** können grundsätzlich alle in UTF-8 definierten Zeichen (insbesondere Buchstaben, Zahlen und Sonderzeichen) enthalten.

5.3 Erläuterung der Variablen

Variable	Typ	Erläuterung
doc_id	String	(Nur CSV) Dateiname der extrahierten TXT-Datei.
text	String	(Nur CSV) Volltext der Entscheidung. Die einzige editorische Änderung gegenüber den TXT-Dateien ist die Zusammenfügung von über Zeilengrenzen gebrochenen Wörtern.
url	String	(Nur CSV) Link zum Volltext der Entscheidung in der Datenbank.

(continued)

Variable	Typ	Erläuterung
gericht	String	Name des Gerichts. Es ist nur der Wert »BGH« vergeben. Dies ist der ECLI-Code für »Bundesgerichtshof«. Diese Variable dient vor allem zur einfachen und transparenten Verbindung der Daten mit anderen Datensätzen.
datum	Date	Datum der Entscheidung im Format YYYY-MM-DD (ISO-8601).
entscheidungsjahr	Integer	(Nur CSV) Jahr der Entscheidung im Format YYYY (ISO-8601).
entscheidung_typ	String	(Nur CSV-Datei) Der Typ der Entscheidung. Es sind die Werte »B« (Beschluss), »U« (Urteil) und »V« (Verfügung) vergeben. Wurde durch <i>regular expressions</i> aus der Variable »text« berechnet.
leitsatz	String	Ob es sich um eine Leitsatzentscheidung handelt. Wenn ja ist der Wert »LE«, ansonsten »NA«. Wurde für jede Entscheidung bejaht, die in der Variable »bemerkung« die Zeichenfolge »Leitsatz« enthält.
spruchkoerper_db	String	Der Spruchkörper wie er in der amtlichen Datenbank des Bundesgerichtshofs eingetragen ist. Gekürzt auf eine für Dateinamen angemessene Länge und konvertiert in arabische Zahlen. Die übrigen Senate haben eindeutige Namen.
spruchkoerper_az	String	Der im Aktenzeichen angegebene Spruchkörper. Arabische Ziffern stehen für die Strafsenate, römische Ziffern für die Zivilsenate. Die Spezialsenate und großen Senate sind in dieser Variable nicht codiert (weil über die Registerzeichen eindeutig identifiziert) und weisen daher immer den Wert »NA« auf. Nutzen Sie für Analysen aller Senate bitte die Variablen »spruchkoerper_db« oder »registerzeichen«.
registerzeichen	String	Das amtliche Registerzeichen. Eine Erläuterung der Abkürzungen findet sich im Abschnitt 6.
verfahrensart	String	Die Verfahrensart, auf die das Registerzeichen hinweist. Siehe auch Abschnitt 6.
eingangsnummer	Integer	Eingangsnummer. Verfahren des gleichen Eingangsjahres erhalten vom Gericht eine fortlaufende Nummer (Ordinalzahl) in der Reihenfolge ihres Eingangs.

(continued)

Variable	Typ	Erläuterung
eingangsjahr_az	Integer	Eingangsjahr laut Aktenzeichen. Das Jahr in dem das Verfahren beim Gericht anhängig wurde. Das Format ist eine zweistellige Jahreszahl (YY).
eingangsjahr_iso	Integer	Eingangsjahr im Format YYYY-MM-DD (ISO-8601).
zusatz_az	String	Der Zusatz zu einem Aktenzeichen, falls vorhanden.
name	String	Der Name der Entscheidung wie er in der amtlichen Datenbank angegeben ist. In den CSV-Dateien ist er identisch mit dem Datenbankeintrag, in den Dateinamen der PDF- und TXT-Dateien wurden bestimmte Sonderzeichen bereinigt um Kompatibilitätsprobleme zu vermeiden.
kollision	Integer	In wenigen Fällen sind am gleichen Tag mehrere Entscheidungen zum gleichen Aktenzeichen ergangen. Diese werden ab der zweiten Entscheidung pro Tag durch eine Kollisions-ID mit einer fortlaufenden Zahl ausdifferenziert. Für die erste Entscheidung ist der Wert der Variable »0«. Die zweite Entscheidung ist mit »1« identifiziert, die dritte mit »2« und so fort. ACHTUNG: diese Variable ist von dem Autor des Datensatzes berechnet und keine amtliche Angabe. Die Nutzung für die Erstellung der ECLI ist daher problematisch.
aktenzeichen	String	Das amtliche Aktenzeichen im Format [senatsnummer] [registerzeichen] [eingangsnummer] / [eingangsjahr] [ggf. zusatz_az]

(continued)

Variable	Typ	Erläuterung
ecli	String	(Nur CSV-Datei) Experimentell! Der European Case Law Identifier (ECLI) der Entscheidung. Jeder Entscheidung ist eine einzigartige ECLI zugewiesen. Die ECLI ist vor allem dann hilfreich, wenn dieser Datensatz mit anderen Datensätzen zusammengeführt und Duplikate vermieden werden sollen. Alle inhaltlichen Bestandteile der ECLI sind in diesem Datensatz zusätzlich auch anderen und besser verständlichen Variablen zugewiesen. Nutzen Sie bevorzugt diese anderen Variablen, statt Informationen aus der ECLI zu extrahieren. Die Variable wurde aus den Variablen »gericht«, »entscheidungsjahr«, »datum«, »entscheidung_typ«, »spruchkoerper_az«, »registerzeichen«, »eingangsnummer«, »eingangsjahr_az« und »kollision« durch den Autor des Datensatzes berechnet. ACHTUNG: Da keine amtlichen Angaben zur Kollisionsvariable vorliegen sind die ECLIs für alle Entscheidungen mit Variable »kollision« größer 0 und ihre verwandten Entscheidungen mit Variable »kollision« gleich 0 mit hoher Wahrscheinlichkeit fehlerhaft. Für die weit überwiegende Anzahl der Entscheidungen gibt es jedoch keine Kollisionen und die ECLIs sollten funktionstüchtig sein. Amtlich vergeben werden jedoch nur ECLIs für Entscheidungen ab dem 1. Januar 2016.
praesi	String	(Nur CSV-Datei) Der Nachname des oder der Präsident:in in dessen/deren Amtszeit das Datum der Entscheidung fällt.
v_praesi	String	(Nur CSV-Datei) Der Nachname des oder der Vize-Präsident:in in dessen/deren Amtszeit das Datum der Entscheidung fällt.
bemerkung	String	Zusätzliche Bemerkungen aus der Datenbank des Bundesgerichtshofs. Enthält in der Regel Informationen dazu ob es sich um eine Leitsatzentscheidung handelt, welchen Eigennamen die Entscheidung trägt und wann welche Pressemitteilung darüber herausgegeben wurde.
bghst	Binary	(Nur CSV-Datei) Ob die Entscheidung in die amtliche Sammlung BGHSt aufgenommen wird. »1« falls ja, »0« falls nein.

(continued)

Variable	Typ	Erläuterung
bghz	Binary	(Nur CSV-Datei) Ob die Entscheidung in die amtliche Sammlung BGHZ aufgenommen wird. »1« falls ja, »0« falls nein.
bghr	Binary	(Nur CSV-Datei) Ob die Entscheidung in die Sammlung BGHR aufgenommen wird. »1« falls ja, »0« falls nein.
nachschlagewerk	Binary	(Nur CSV-Datei) Ob die Entscheidung in das Nachschlagewerk aufgenommen wird. »1« falls ja, »0« falls nein.
berichtigung	Logical	(Nur CSV-Datei) Eine Angabe ob es sich um einen Berichtigungsbeschluss handelt, TRUE wenn ja, sonst FALSE. Wurde durch den Autor des Datensatzes aus der Variable »bemerkung« extrahiert.
zeichen	Integer	(Nur CSV-Datei) Die Anzahl Zeichen eines Dokumentes.
tokens	Integer	(Nur CSV-Datei) Die Anzahl Tokens (beliebige Zeichenfolge getrennt durch whitespace) eines Dokumentes. Diese Zahl kann je nach Tokenizer und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Tokenisierung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.
typen	Integer	(Nur CSV-Datei) Die Anzahl <i>einzigartiger</i> Tokens (beliebige Zeichenfolge getrennt durch whitespace) eines Dokumentes. Diese Zahl kann je nach Tokenizer und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Tokenisierung und Typenzählung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.

(continued)

Variable	Typ	Erläuterung
saetze	Integer	(Nur CSV-Datei) Die Anzahl Sätze. Die Definition entspricht in etwa dem üblichen Verständnis eines Satzes. Die Regeln für die Bestimmung von Satzanfang und Satzende sind im Detail allerdings sehr komplex und in »Unicode Standard: Annex No 29« beschrieben. Diese Zahl kann je nach Software und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Zählung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.
version	Date	(Nur CSV-Datei) Die Versionsnummer des Datensatzes im Format YYYY-MM-DD (Langform nach ISO-8601). Die Versionsnummer entspricht immer dem Datum an dem der Datensatz erstellt und die Daten von der Webseite des Gerichts abgerufen wurden.
doi_concept	String	(Nur CSV-Datei) Der Digital Object Identifier (DOI) des Gesamtkonzeptes des Datensatzes. Dieser ist langzeit-stabil (persistent). Über diese DOI kann via www.doi.org immer die aktuellste Version des Datensatzes abgerufen werden. Prinzip F1 der FAIR-Data Prinzipien (»data are assigned globally unique and persistent identifiers«) empfiehlt die Dokumentation jeder Messung mit einem persistenten Identifikator. Selbst wenn die CSV-Dateien ohne Kontext weitergegeben werden kann ihre Herkunft so immer zweifelsfrei und maschinenlesbar bestimmt werden.
doi_version	String	(Nur CSV-Datei) Der Digital Object Identifier (DOI) der konkreten Version des Datensatzes. Dieser ist langzeit-stabil (persistent). Über diese DOI kann via www.doi.org immer diese konkrete Version des Datensatzes abgerufen werden. Prinzip F1 der FAIR-Data Prinzipien (»data are assigned globally unique and persistent identifiers«) empfiehlt die Dokumentation jeder Messung mit einem persistenten Identifikator. Selbst wenn die CSV-Dateien ohne Kontext weitergegeben werden kann ihre Herkunft so immer zweifelsfrei und maschinenlesbar bestimmt werden.

(continued)

Variable	Typ	Erläuterung
lizenz	String	Die Lizenz für den Gesamtdatensatz. In diesem Datensatz immer »Creative Commons Zero 1.0 Universal«.

6 Registerzeichen

Die Registerzeichen des BGH sind in der *Anweisung für die Verwaltung des Schriftguts in Rechtssachen bei der Geschäftsstelle des Bundesgerichtshofes und des Generalbundesanwalts beim Bundesgerichtshof - Aktenordnung Bundesgerichtshof* definiert. Die im Datensatz enthaltenen Registerzeichen wurden jeweils um die runden Klammern bereinigt, um Probleme bei der Nutzung unter Windows zu vermeiden.

Die Tabelle der Registerzeichen und der ihnen zugeordneten Verfahrensarten stammt aus dem folgenden Datensatz: »Seán Fobbe (2021). Aktenzeichen der Bundesrepublik Deutschland (AZ-BRD). Version 1.0.1. Zenodo. DOI: 10.5281/zenodo.4569564.«

Registerzeichen	Verfahrensart
AK	Aktenkontrolle für Haftprüfungsverfahren
AR	Allgemeines Register
AR(Anw)	Allgemeines Register (Anwaltssachen)
AR(Not)	Allgemeines Register (Notarsachen)
AR(Ri)	Allgemeines Register (Dienstgericht des Bundes)
AR(VS)	Entscheidungen über Justizverwaltungsakte (Strafsachen)
AR(VZ)	Entscheidungen über Justizverwaltungsakte (Zivilsachen)
AR(Vollz)	Vorlegungssachen (Strafvollzugssachen)
ARZ	Allgemeines Register und Gerichtsstandsbestimmungen (Zivilsachen)
ARs	Allgemeines Register und Gerichtsstandsbestimmungen (Strafsachen)
AnwSt	Anträge auf Einleitung des anwaltsgerichtlichen Verfahrens und Beschwerden in Ordnungsmittelverfahren gegen Rechtsanwälte beim Bundesgerichtshof
AnwSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden gegen Entscheidungen eines Anwaltsgerichtshofes
AnwSt(R)	Revisionen gegen Urteile eines Anwaltsgerichtshofes
AnwZ	Klagen über Entscheidungen in Zulassungssachen oder gegen sonstige Verwaltungsakte betreffend Rechtsanwälte beim Bundesgerichtshof
AnwZ(B)	Beschwerden gegen Entscheidungen eines Anwaltsgerichtshofes
AnwZ(Brfg)	Berufungen und Anträge auf Zulassung der Berufung gegen Entscheidungen eines Anwaltsgerichtshofes
AnwZ(P)	Klagen über die Anfechtung von Wahlen und Beschlüssen der Rechtsanwaltskammer beim BGH und der Bundesrechtsanwaltskammer
BGs	Einzelne richterliche Anordnungen des Ermittlungsrichters (Strafsachen)

BLw	Rechtsbeschwerden in Landwirtschaftssachen der freiwilligen Gerichtsbarkeit
EnRB	Rechtsbeschwerden in Bußgeldverfahren nach dem EnWG
EnVR	Rechtsbeschwerden in energiewirtschaftsrechtlichen Verwaltungssachen nach dem EnWG
EnVZ	Nichtzulassungsbeschwerden in energiewirtschaftsrechtlichen Verwaltungssachen nach dem EnWG
EnZA	Anträge außerhalb eines in der Rechtsmittelinstanz anhängigen Rechtsstreits in bürgerlichen Rechtsstreitigkeiten nach dem EnWG
EnZB	Rechtsbeschwerden und Beschwerden in bürgerlichen Rechtsstreitigkeiten nach dem EnWG
EnZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision und Anträge auf Zulassung der Sprungrevision in bürgerlichen Rechtsstreitigkeiten nach dem EnWG
GSSt	Großer Senat (Strafsachen)
GSZ	Großer Senat (Zivilsachen)
GmSOGB	Gemeinsamer Senat der Obersten Gerichtshöfe des Bundes
KRB	Rechtsbeschwerden in Kartellbußgeldverfahren
KVR	Rechtsbeschwerden (Kartellverwaltungssachen)
KVZ	Nichtzulassungsbeschwerden (Kartellverwaltungssachen)
KZA	Anträge außerhalb eines in der Revisionsinstanz anhängigen Rechtsstreits (Kartellsachen)
KZB	Rechtsbeschwerden und Beschwerden in bürgerlichen Rechtsstreitigkeiten (Kartellsachen)
KZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision und Anträge auf Zulassung der Sprungrevision in bürgerlichen Rechtsstreitigkeiten (Kartellsachen)
LwZA	Anträge außerhalb eines in der Revisionsinstanz für Landwirtschaftssachen anhängigen Verfahrens
LwZB	Beschwerden und Rechtsbeschwerden in Landwirtschaftssachen der streitigen bürgerlichen Gerichtsbarkeit
LwZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision und Anträge auf Zulassung der Sprungrevision (Landwirtschaftssachen)
NotSt(B)	Beschwerden gegen Beschlüsse der Oberlandesgerichte in Disziplinarsachen gegen Notare
NotSt(Brfg)	Berufungen und Anträge auf Zulassung der Berufung gegen Urteile der Oberlandesgerichte in Disziplinarsachen gegen Notare
NotZ	Verwaltungsstreitverfahren in Notarsachen und Beschwerden gegen Entscheidungen der Oberlandesgerichte
NotZ(Brfg)	Berufungen und Anträge auf Zulassung der Berufung gegen Entscheidungen der Oberlandesgerichte in Notarsachen

PatAnwSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden nach der PatAO
PatAnwSt(R)	Revisionen nach der PatAO
PatAnwZ	Verwaltungsstreitverfahren in Patentanwaltssachen; Berufungen und Anträge auf Zulassung der Berufung und Beschwerden gegen Entscheidungen der Oberlandesgerichte
RiSt	Disziplinarsachen und Anträge auf Einleitung oder Einstellung des förmlichen Disziplinarverfahrens, auf vorläufige Dienstenthebung, auf Einbehaltung von Dienstbezügen sowie auf Aufhebung dieser Maßnahmen gegen Richter und Staatsanwälte im Bundesdienst sowie Mitglieder des Bundesrechnungshofes
RiSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden der Richter, Staatsanwälte und Mitglieder des Bundesrechnungshofes gegen Disziplinarverfügungen
RiSt(R)	Revisionen in Disziplinarsachen nach dem Deutschen Richtergesetz
RiZ	Anträge betreffend Richter im Bundesdienst und Mitglieder des Bundesrechnungshofes auf gerichtliche Entscheidung im Versetzungs- und Prüfungsverfahren sowie auf vorläufige Untersagung der Amtsgeschäfte
RiZ(B)	Beschwerden gegen die Nichtzulassung der Revision nach dem Deutschen Richtergesetz
RiZ(R)	Revisionen in Versetzungs- und Prüfungsverfahren nach dem Deutschen Richtergesetz
StB	Beschwerden (Strafsachen)
StR	Revisionen und Vorlegungssachen nach § 121 Abs.1 Nr.1, Abs. 2 GVG, § 79 Abs. 3 OWiG, §§ 13 Abs. 4, 25 StrRehaG (Strafsachen)
StbSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden in berufsgerichtlichen Verfahren (Steuerberater- und Steuerbevollmächtigten-sachen)
StbSt(R)	Revisionen in berufsgerichtlichen Verfahren (Steuerberater- und Steuerbevollmächtigten-sachen)
VGS	Vereinigte Große Senate
VRG	Ersuchen des Bundesverfassungsgerichts zu Vorlagen nach Art. 100 GG und zu Verfassungsbeschwerden
WpSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden in berufsgerichtlichen Verfahren (Wirtschaftsprüfersachen)
WpSt(R)	Revisionen in berufsgerichtlichen Verfahren (Wirtschaftsprüfersachen)
ZA	Anträge außerhalb eines in der Rechtsmittelinstanz anhängigen Verfahrens (Zivilsachen)
ZB	Beschwerden, Rechtsbeschwerden, weitere Beschwerden, Beschwerden gegen die Nichtzulassung der Revision nach dem BEG (Zivilsachen)

ZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision, Anträge auf Zulassung der Sprungrevision, Berufungen in Patent-sachen (Zivilsachen)
ZR(Ü)	Erstinstanzliche Klagen auf Entschädigung wegen überlanger Ge-richtsverfahren und strafrechtlicher Ermittlungsverfahren (Zivilsa-chen)

7 Präsident:innen

7.1 Hinweise

- Die Personendaten stammen aus folgendem Datensatz: »Seán Fobbe and Tilko Swalve (2021). Presidents and Vice-Presidents of the Federal Courts of Germany (PVP-FCG). Version 2021-04-08. Zenodo. DOI: 10.5281/zenodo.4568682«.
- Das Datum bezieht sich jeweils auf das Amt als Präsident:in, nicht auf die Amtszeit als Richter:in.

7.2 Lebensdaten

Nachname	Vorname	Amtsantritt	Amtsende	Geboren	Gestorben
Weinkauff	Hermann	1950-10-01	1960-03-31	1894-02-10	1981-07-09
Heusinger	Bruno	1960-04-01	1968-03-31	1900-03-02	1987-08-03
Fischer	Robert	1968-04-01	1977-09-30	1911-08-22	1983-03-04
Pfeiffer	Gerd	1977-10-01	1987-12-31	1919-12-22	2007-02-01
Odersky	Walter	1988-01-01	1996-07-31	1931-07-17	NA
Geiss	Karlmann	1996-08-01	2000-05-31	1935-05-31	NA
VACANCY-1	VACANCY-1	2000-06-01	2000-07-14	NA	NA
Hirsch	Günter	2000-07-15	2008-01-31	1943-01-30	NA
Tolksdorf	Klaus	2008-02-01	2014-01-31	1948-11-14	NA
VACANCY-2	VACANCY-2	2014-02-01	2014-06-30	NA	NA
Limberg	Bettina	2014-07-01	NA	1960-04-05	NA

7.3 Dienstalter und Lebensalter

Nachname	Vorname	Alter (Amtsantritt)	Alter (Amtsende)	Alter (Tod)
Weinkauff	Hermann	56	66	87
Heusinger	Bruno	60	68	87
Fischer	Robert	56	66	71
Pfeiffer	Gerd	57	68	87
Odersky	Walter	56	65	NA
Geiss	Karlmann	61	65	NA
Hirsch	Günter	57	65	NA
Tolksdorf	Klaus	59	65	NA
Limberg	Bettina	54	NA	NA

8 Vize-Präsident:innen

8.1 Hinweise

- Die Personendaten stammen aus folgendem Datensatz: »Seán Fobbe and Tilko Swalve (2021). Presidents and Vice-Presidents of the Federal Courts of Germany (PVP-FCG). Version 2021-04-08. Zenodo. DOI: 10.5281/zenodo.4568682«.
- Das Datum bezieht sich jeweils auf das Amt als Vize-Präsident:in, nicht auf die Amtszeit als Richter:in.

8.2 Lebensdaten

Nachname	Vorname	Amtsantritt	Amtsende	Geboren	Gestorben
Glanzmann	Roderich	1965-05-17	1972-04-30	1904-04-15	1988-08-07
VACANCY-1	VACANCY-1	1972-05-01	1972-05-22	NA	NA
Hauß	Fritz	1972-05-23	1976-10-31	1908-10-27	2003-04-30
VACANCY-2	VACANCY-2	1976-11-01	1976-11-02	NA	NA
Pfeiffer	Gerd	1976-11-03	1977-09-30	1919-12-22	2007-02-01
Stimpel	Walter	1977-10-01	1985-11-30	1917-11-29	2008-01-13
VACANCY-3	VACANCY-3	1985-12-01	1985-12-01	NA	NA
Thumm	Ludwig	1985-12-02	1988-04-30	1920-04-13	2011-04-11
Salger	Hannskarl	1988-05-01	1994-11-30	1929-11-02	2010-09-17
Hagen	Horst	1994-12-01	1999-02-28	1934-01-05	2019-12-05
Jähnke	Burkhard	1999-03-01	2002-05-31	1937-05-14	NA
Wenzel	Joachim	2002-06-01	2005-06-30	1940-06-23	2009-08-29
Müller	Gerda	2005-07-01	2009-06-30	1944-06-26	NA
Schlick	Wolfgang	2009-07-01	2015-07-31	1950-03-29	NA
VACANCY-4	VACANCY-4	2015-08-01	2016-12-01	NA	NA
Ellenberger	Jürgen	2016-12-02	NA	1960-04-19	NA

8.3 Dienstalter und Lebensalter

Nachname	Vorname	Alter (Amtesantritt)	Alter (Amtesende)	Alter (Tod)
Glanzmann	Roderich	61	68	84
Hauß	Fritz	63	68	94
Pfeiffer	Gerd	56	57	87
Stimpel	Walter	59	68	90
Thumm	Ludwig	65	68	90
Salger	Hannskarl	58	65	80
Hagen	Horst	60	65	85
Jähnke	Burkhard	61	65	NA
Wenzel	Joachim	61	65	69
Müller	Gerda	61	65	NA
Schlick	Wolfgang	59	65	NA
Ellenberger	Jürgen	56	NA	NA

9 Linguistische Kennzahlen

9.1 Erläuterung der Kennzahlen und Diagramme

Zur besseren Einschätzung des inhaltlichen Umfangs des Korpus dokumentiere ich an dieser Stelle die Verteilung der Werte für einige klassische linguistische Kennzahlen:

Kennzahl	Definition
Zeichen	Zeichen entsprechen grob den <i>Graphemen</i> , den kleinsten funktionalen Einheiten in einem Schriftsystem. Beispiel: das Wort »RichterIn« besteht aus 9 Zeichen.
Tokens	Eine beliebige Zeichenfolge, getrennt durch whitespace-Zeichen, d.h. ein Token entspricht in der Regel einem »Wort«, kann aber auch Zahlen, Sonderzeichen oder sinnlose Zeichenfolgen enthalten, weil es rein syntaktisch berechnet wird.
Typen	Einzigartige Tokens. Beispiel: wenn das Token »Verfassungsrecht« zehnmal in einer Entscheidung vorhanden ist, wird es als ein Typ gezählt.
Sätze	Entsprechen in etwa dem üblichen Verständnis eines Satzes. Die Regeln für die Bestimmung von Satzanfang und Satzende sind im Detail aber sehr komplex und in »Unicode Standard: Annex No 29« beschrieben.

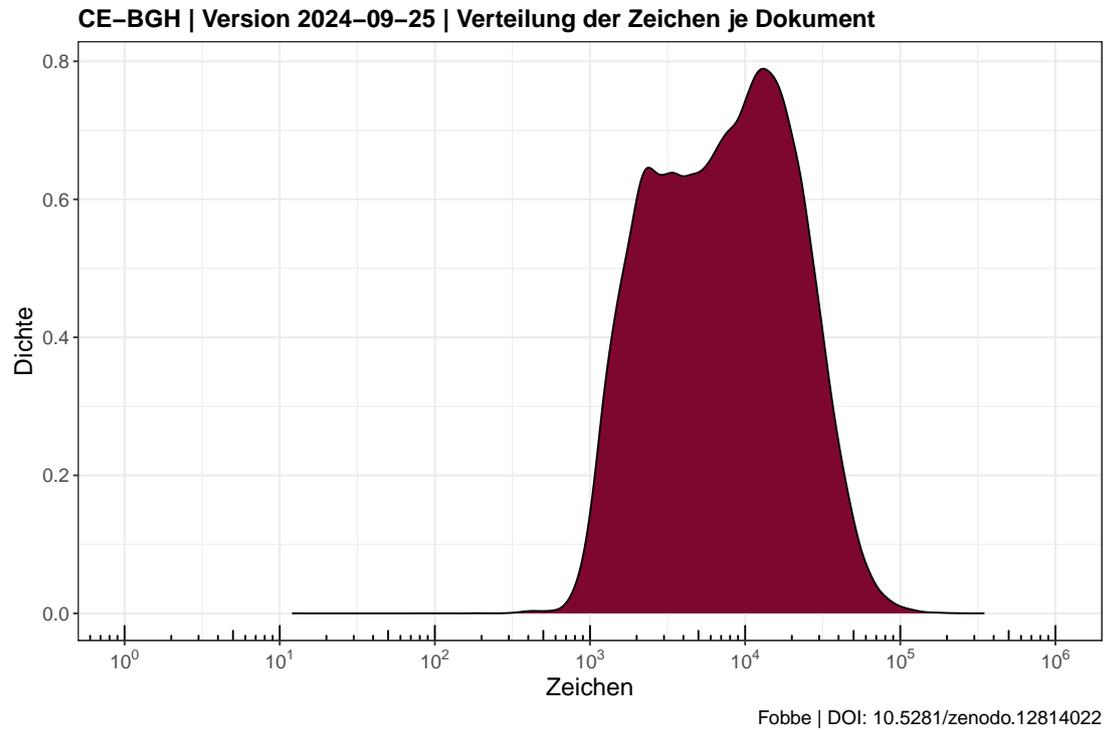
Es handelt sich bei den Diagrammen jeweils um »Density Charts«, die sich besonders dafür eignen die Schwerpunkte von Variablen mit stark schwankenden numerischen Werten zu visualisieren. Die Interpretation ist denkbar einfach: je höher die Kurve, desto dichter sind in diesem Bereich die Werte der Variable. Der Wert der y-Achse kann außer Acht gelassen werden, wichtig sind nur die relativen Flächenverhältnisse und die x-Achse.

Vorsicht bei der Interpretation: Die x-Achse ist logarithmisch skaliert, d.h. in 10er-Potenzen und damit nicht-linear. Die kleinen Achsen-Markierungen zwischen den Schritten der Exponenten sind eine visuelle Hilfestellung um diese nicht-Linearität zu verstehen.

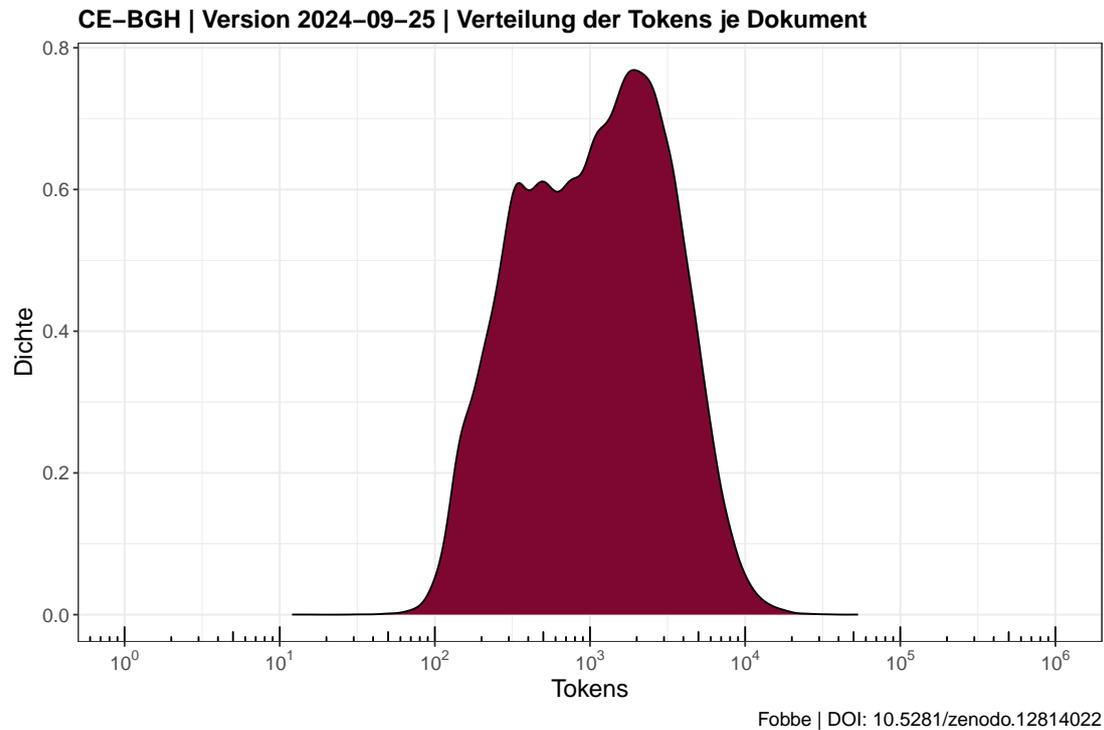
9.2 Werte der Kennzahlen

Variable	Summe	Min	Quart1	Median	Mittel	Quart3	Max
zeichen	906,902,015	12	3,078	7,486.5	11,643.07	16,054	348,611
tokens	137,780,013	0	443	1,120.0	1,768.86	2,442	53,485
typen	746,077	0	215	435.0	542.93	766	5,582
saetze	7,270,257	0	29	64.0	93.34	127	2,700

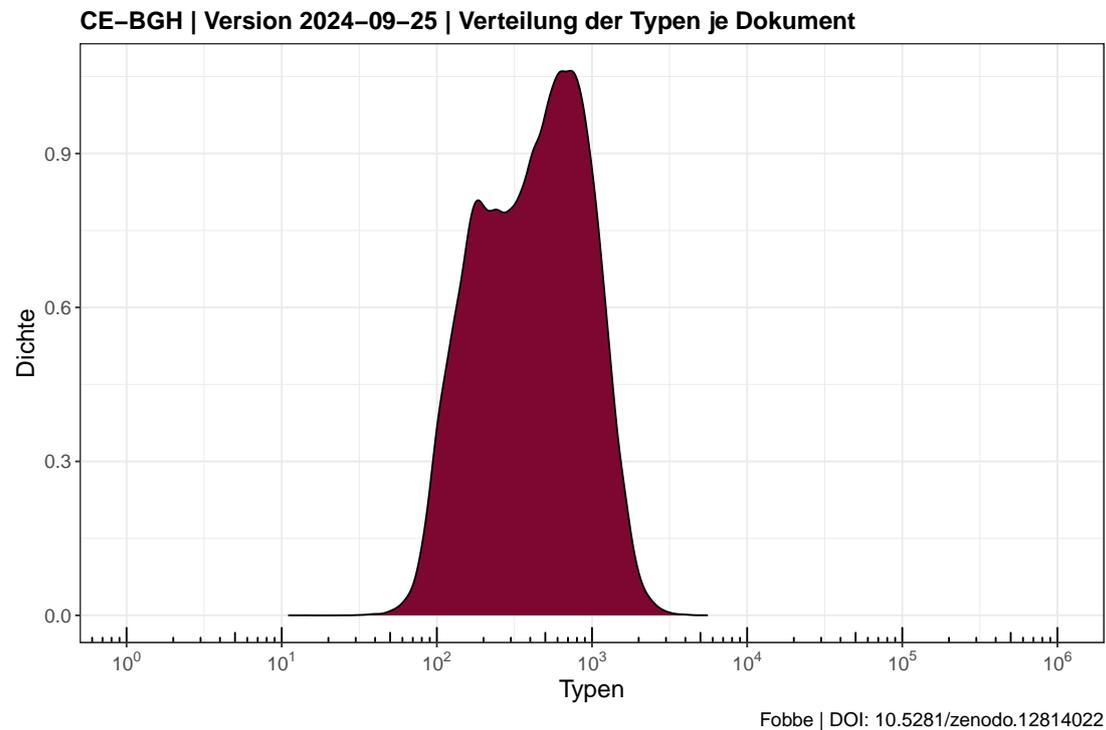
9.3 Verteilung Zeichen



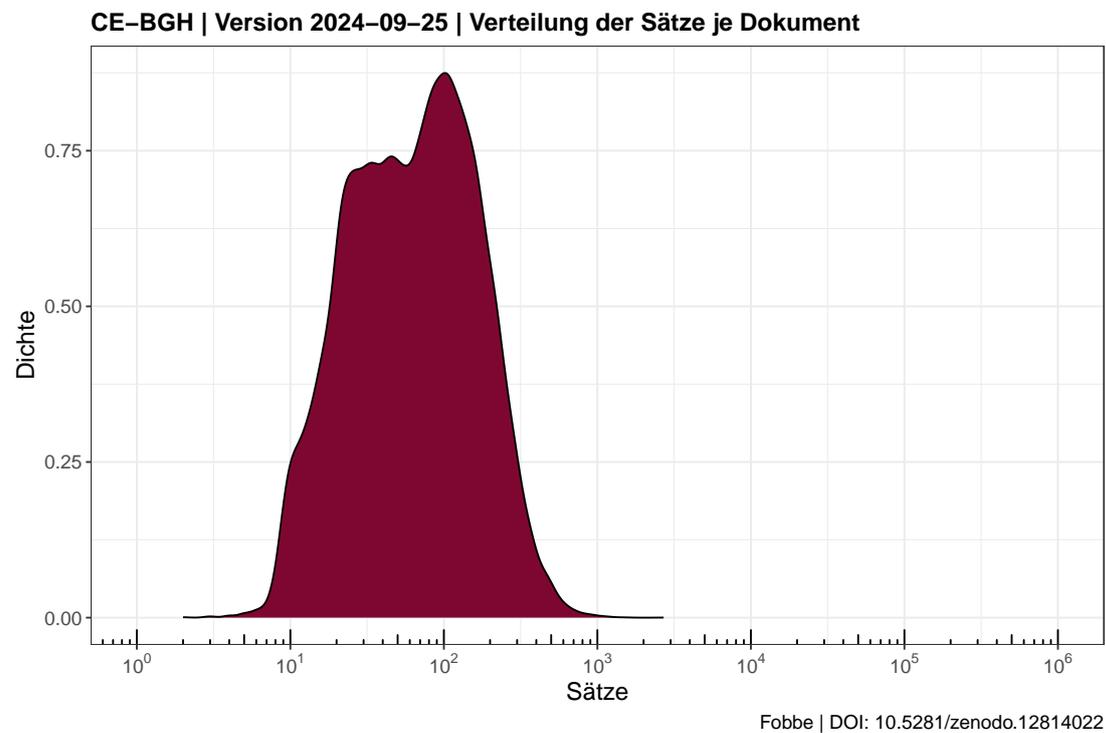
9.4 Verteilung Tokens



9.5 Verteilung Typen



9.6 Verteilung Sätze



10 Zitationsnetzwerk des Bundesgerichtshofs (Beta)

10.1 Überblick

Der Datensatz enthält zusätzlich eine spezialisierte Variante, die Zitate des BGH aus den Entscheidungstexten zu seiner eigenen Rechtsprechung extrahiert und in strukturierter Form aufbereitet.

Diese Variante ist noch in der Beta-Testphase. Folgende Rechtsprechungszitate sind enthalten:

- Zitate von Aktenzeichen zu Aktenzeichen
- Zitate von Aktenzeichen zu BGHZ
- Zitate von Aktenzeichen zu BGHSt

Achtung: Zitierende Entscheidungen können nur solche sein, für die im Korpus der Volltext dokumentiert ist (d.h. normalerweise ab 2000). Zitierte Entscheidungen können aus der gesamten Rechtsprechung stammen (d.h. aus allen Jahren).

Zitate unter Angabe des Aktenzeichens sind weniger genau als Zitate zu konkreten Entscheidungen (bei denen Datum und ggf. Kollisionsziffer nötig sind). Sie stellen aber eine gute Näherung dar. Die Auflösung der Zitate auf Entscheidungsebene ist geplant und wird in Zukunft mitveröffentlicht.

Eingehende BGHZ- und BGHSt-Zitate können exakt einer eingehenden Entscheidung zugeordnet werden. Die Quell-Dokumente sind jedoch nur mit dem Aktenzeichen hinterlegt, um eine Konkordanz mit dem Rest des Datensatzes herzustellen. Die Auflösung von Quell-Dokumenten nach BGHZ/BGHSt ist mir leider nicht möglich, da ich aktuell keine Entsprechungstabelle zwischen amtlichen Sammlungen und Aktenzeichen/Datum-Zitaten habe.

10.2 Technische Hinweise

Das Zitationsnetzwerk wird als GraphML-Datei angeboten und kann z.B. einfach in graphische Software wie Gephi⁸ importiert und ohne Programmierkenntnisse genutzt werden. Formal handelt sich um einen gewichteten, gerichteten Graphen (Digraph). Die Anzahl der Knoten gibt die Anzahl der BGHZ-Entscheidungen, BGHSt-Entscheidungen und Aktenzeichen mit eingehenden und/oder ausgehenden Zitaten an. Die Anzahl der Kanten gibt die Anzahl der Knoten-Paare mit mindestens einem Zitat an. Die Gewichte der Kanten geben die Anzahl der Zitate zwischen Knoten an. Die Ausgangsstärke gibt die Summe aller einfachen Zitate an.

Beachten Sie auch bitte folgende Punkte:

- Das gesamte Netzwerk ist sehr groß und die Analyse ist daher ohne weitere Einschränkungen rechenintensiv und anspruchsvoll. In der Regel sollten Sie das Netzwerk auf die für Sie interessanten Teile reduzieren.
- Zur Reduktion des Netzwerks auf eine handliche Größe stelle ich zwei Variablen bereit: den Senat und das Registerzeichen. Sie könnten beispielsweise über das Registerzeichen nur die Strafsenate betrachten oder über die Senatsangabe nur den Zivilsenat mit der für Sie interessanten Spezialmaterie.

⁸ <https://gephi.github.io/>

- Die Extraktion mit *regular expressions* ist nicht perfekt. Es kann daher sein, dass Zitate fehlen, wenn sie nicht als solche erkannt wurden, wegen Tippfehlern, ungewöhnlichem Textumfeld etc. Es ist aktuell unklar wieviele Zitate fehlen könnten, weil es keinen Goldstandard zum Abgleich gibt. Wenn Ihnen größere Fehlbestände auffallen, melden Sie sich bitte.

Metric	Value	Metric	Value
Number of Nodes	101,474.00	Mean In-Degree	4.35
Number of Edges	441,884.00	Max In-Degree	559.00
Strength (Out)	593,154.00	Min In-Degree	0.00
Mean Degree	8.71	Mean Out-Degree	4.35
Max Degree	559.00	Max Out-Degree	116.00
Min Degree	0.00	Min Out-Degree	0.00

10.3 Methodik Aktenzeichen

Dieser Datensatz enthält sowohl zitierte Aktenzeichen (Aktenzeichen-zu-Aktenzeichen-Zitate), als auch Zitate von Aktenzeichen zu BGHZ- und BGHSt-Entscheidungen (Aktenzeichen-zu-Sammlung-Zitate).

Aktenzeichen sind verhältnismäßig einfach zu erfassen. Bei den Aktenzeichen des BGH aber mit der Besonderheit, dass die Struktur zwischen nummerierten Senate (Zivil- und Strafsenate) und Spezialsenaten abweicht. Bei nummerierten Senaten steht am Anfang des Aktenzeichens die Senatsnummer (z.B. XIII ZB 21/20). Bei Spezialsenaten entfällt diese und das Registerzeichen identifiziert den Senat zweifelsfrei (z.B. AnwZ (B) 12/99). Die Funktion *f.citation_network.R* erstellt daher für jeden der beiden Fälle in mehreren Schritten eine komplexe REGEX, die jeweils die relevanten Registerzeichen in die Suche aufnimmt und arabische/römische Ziffern bei den nummerierten Senaten voranstellt. Der Source Code ist zu komplex um ihn hier im Detail zu besprechen, sehen Sie sich bei Interesse bitte die Funktion genauer an.

Um konkrete Entscheidungen zu zitieren müsste zusätzlich zum Aktenzeichen noch das Datum berücksichtigt werden. Weil dies die REGEX deutlich komplizierter macht, ist dieser Schritt noch in Arbeit. Im Datensatz sind allerdings 91.78 % aller ausgehenden Aktenzeichen einzigartig (unabhängig vom Datum), sodass das Aktenzeichen eine gute Näherung darstellt.

10.4 Methodik BGHZ und BGHSt

Die Zitate zu den amtlichen Sammlungen BGHZ und BGHSt werden aus dem Volltext in einem Zwei-Stufen-Verfahren extrahiert, ähnlich wie in Coupette, *Juristische Netzwerkforschung* (Mohr Siebeck 2019), S. 241–244.

10.4.1 Erste Stufe

In der **ersten Stufe** werden die Zitierblöcke lokalisiert und aus dem Volltext gesammelt. Es wird die starke Annahme getroffen, dass Zitierblöcke mit »BGHZ« oder »BGHSt«

(ignoriert Groß- und Kleinschreibung) eingeleitet werden und nur Whitespace, Zahlen, gewisse Sonderzeichen und gewisse Buchstaben enthalten.

Zitierblöcke enden in der Regel mit einer runden Klammer, die in der REGEX nicht enthalten ist, um sie als Grenzzeichen zu nutzen. Auch Gleichheitszeichen (=) sind nicht enthalten, damit die REGEX vor einem Hinweis auf einen alternative Abdruck abbricht.

Die konkreten regular expressions (REGEX) sind die folgenden:

```
"BGHZ[\\s\\d\\[\\];,\\.<>Rnfu-]+" # BGHZ
"BGHSt[\\s\\d\\[\\];,\\.<>Rnfu-]+" # BGHSt
```

Ein gefundener Zitierblock sieht beispielsweise so aus:

BGHZ 133, 254, 259 ff.; 152, 331, 337; 159, 280, 287 f.; 167, 252, 256

10.4.2 Zweite Stufe

In der **zweiten Stufe** werden aus allen Zitierblöcken die einzelnen Zitate extrahiert, standardisiert und mit der Ausgangsentscheidung verbunden. Die Extraktion trifft die starke Annahme, dass eine Entscheidung der amtlichen Sammlungen entweder mit »BGHZ« bzw. »BGHSt« oder bei einem Mehrfachzitat in einem Zitierblock mit einem Semikolon eingeleitet wird. Folgende REGEX kommen dabei zum Einsatz:

```
# BGHZ
regex.cite <- paste0("(BGHZ|;)\\s*", # hooks
                    "\\d{1,3},\\s*", # Volume
                    "\\d{1,3}") # Page

print(regex.cite)
```

```
## [1] "(BGHZ|;)\\s*\\d{1,3},\\s*\\d{1,3}"
```

```
# BGHSt
regex.cite <- paste0("(BGHSt|;)\\s*", # hooks
                    "\\d{1,3},\\s*", # Volume
                    "\\d{1,3}") # Page

print(regex.cite)
```

```
## [1] "(BGHSt|;)\\s*\\d{1,3},\\s*\\d{1,3}"
```

Damit findet man zwei Varianten von Einzelzitaten:

- »BGHZ 133, 254«
- »; 152, 331«

Die Einzelzitate werden anschließend bereinigt und standardisiert. Zum Ende hin werden Selbstzitate entfernt und Metadaten hinzugefügt.

10.4.3 Grenzen

Die Extraktion mit regulären Ausdrücken hat Grenzen. Insbesondere folgende Probleme führen zur Nichterkennung von Zitaten:

- Tippfehler (außer Groß- und Kleinschreibung)
- Unregelmäßige Zitierweise
- Verkürzte Schreibweisen, beispielsweise BVerfGE 60, 162: »BVerfGE 3, 19 (27), 383 (394); 4, 375 (381 f.);« — das Beispiel stammt aus Coupette (2019: 246)
- Einfügung von Entscheidungsnamen wie in BVerfGE 42, 143: »BVerfGE 7, 198 (205ff) - Lüth -; 18, 85 (92f); 30, 173 (187f, 196f) - Mephisto -; 32, 311 (316)« — das Beispiel stammt aus Coupette (2019: 246)

10.5 Visualisierungen des Zitationsnetzwerks

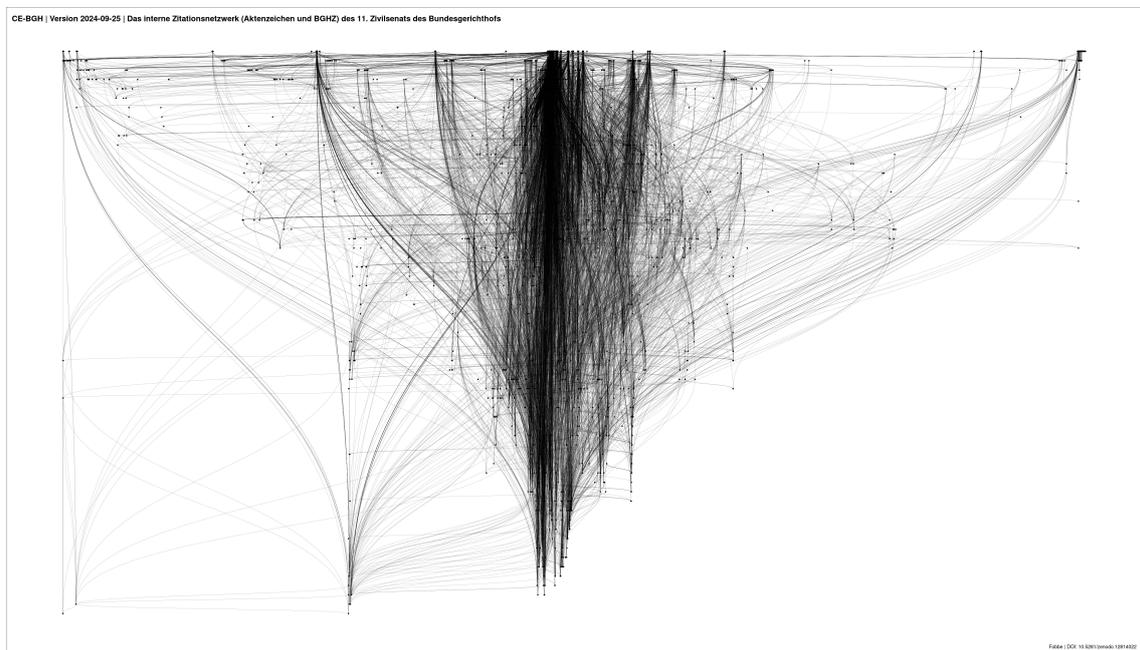
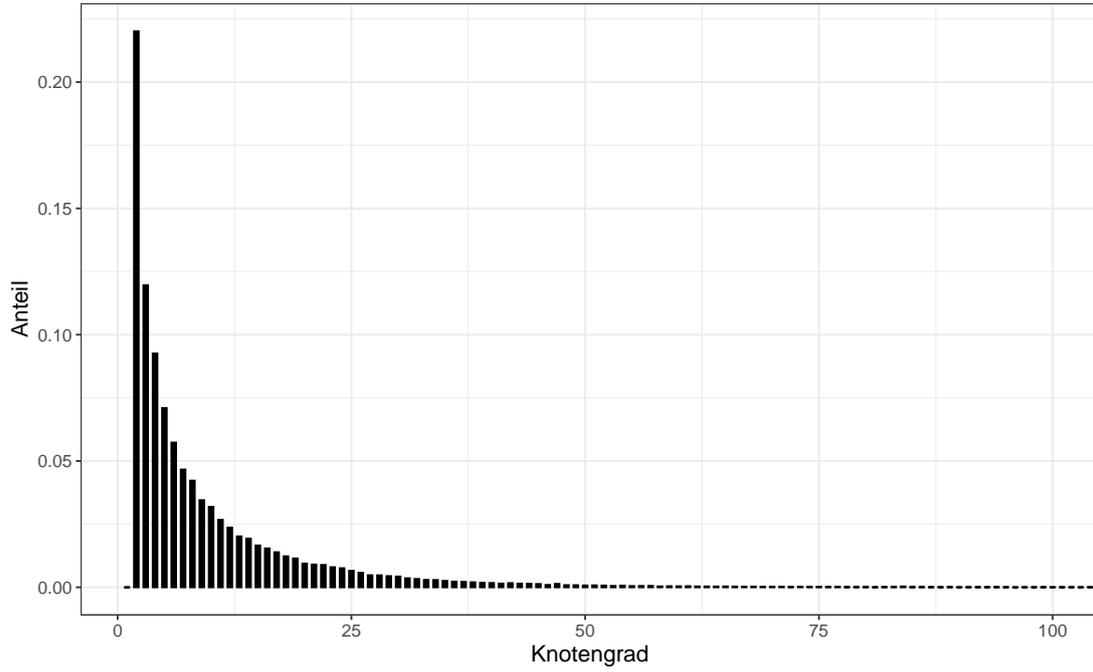


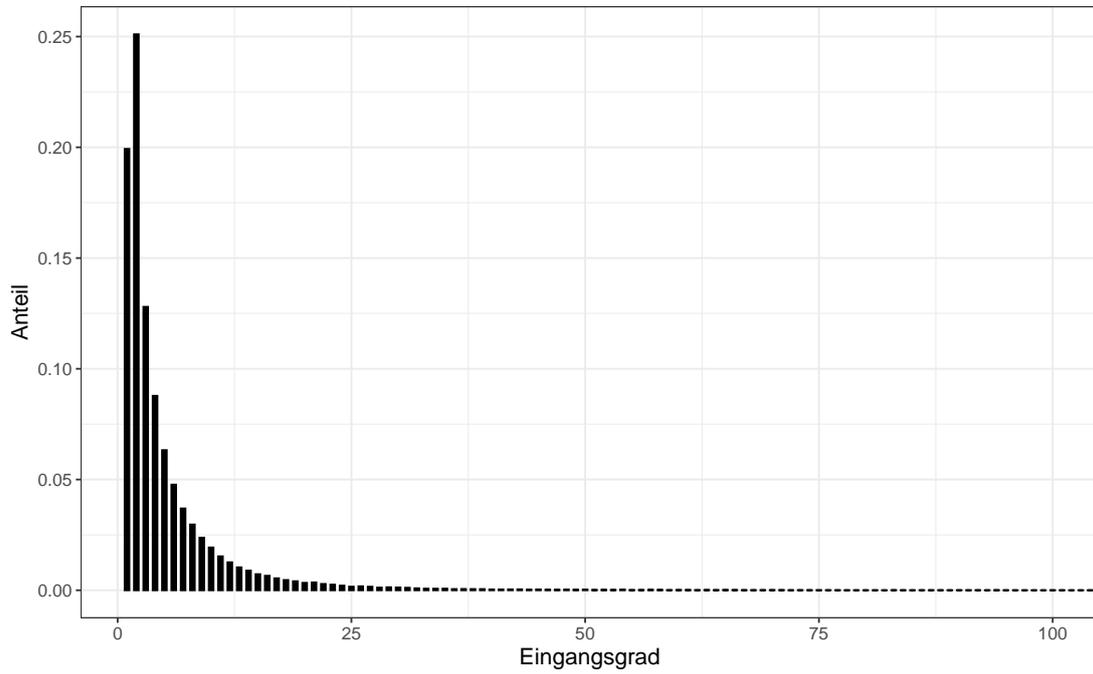
Abbildung 2: Das interne Zitationsnetzwerk (AktENZEICHEN) des 11. Zivilsenats des Bundesgerichtshofs (Sugiyama Layout).

CE-BGH | Version 2024-09-25 | Verteilung der Knotengrade bis 100 im Zitationsnetzwerk



Fobbe | DOI: 10.5281/zenodo.12814022

CE-BGH | Version 2024-09-25 | Verteilung der Eingangsgrade bis 100 im Zitationsnetzwerk



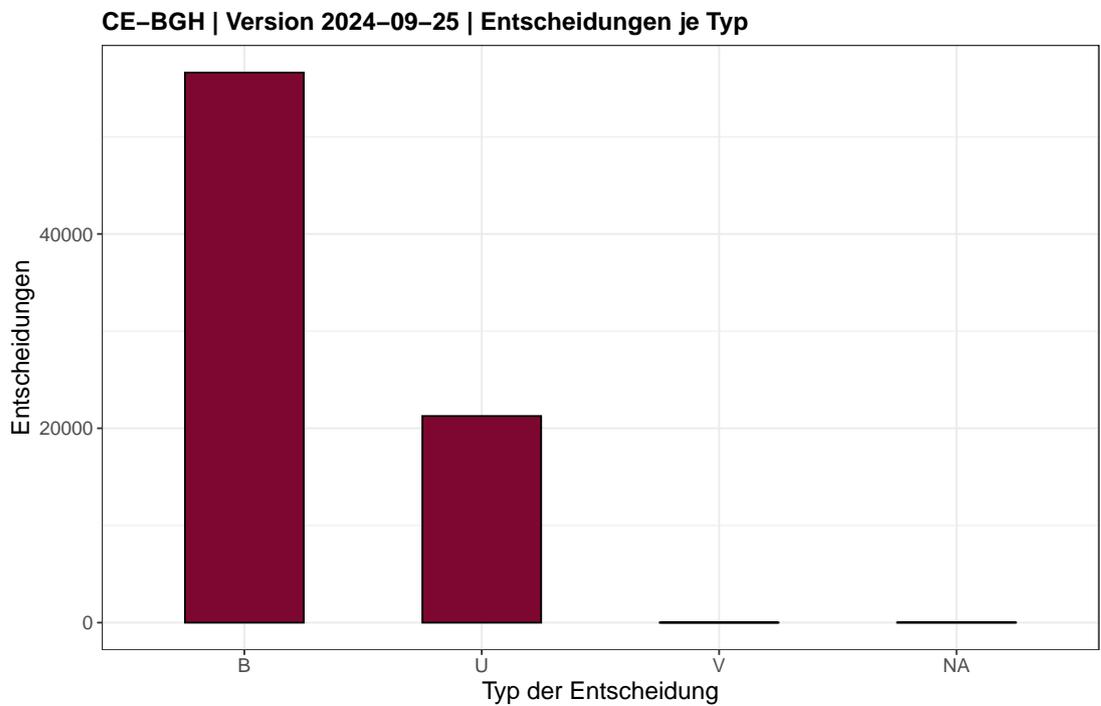
Fobbe | DOI: 10.5281/zenodo.12814022

11 Inhalt des Korpus

11.1 Zusammenfassung

Variable	Anzahl	Min	Quart1	Median	Mean	Quart3	Max
entscheidungsjahr	25	2000	2006	2012	2012.21	2018	2024
eingangsjahr_iso	36	1980	2005	2011	2011.35	2017	2024
eingangsnummer	1076	1	55	156	202.66	307	1743

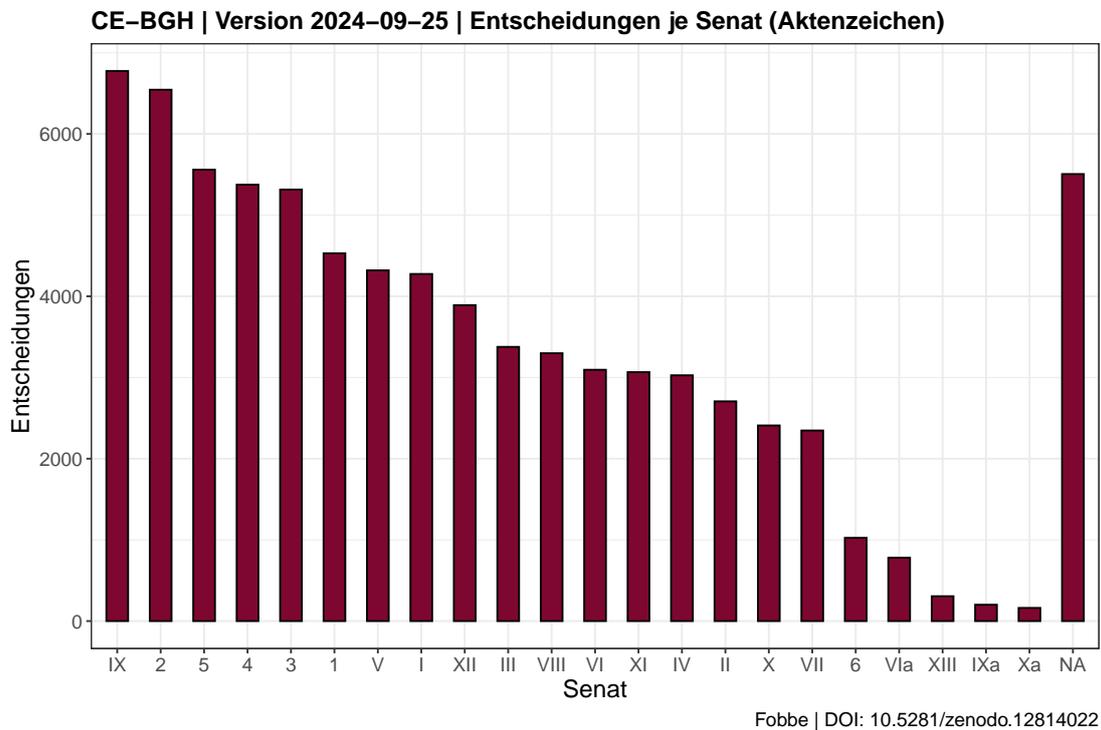
11.2 Nach Typ der Entscheidung



Fobbe | DOI: 10.5281/zenodo.12814022

Typ	Entscheidungen	% Gesamt	% Kumulativ
NA	20	0.03	0.03
B	56604	72.67	72.70
U	21259	27.29	99.99
V	9	0.01	100.00
Total	77892	100.00	100.00

11.3 Nach Spruchkörper (Aktenzeichen)

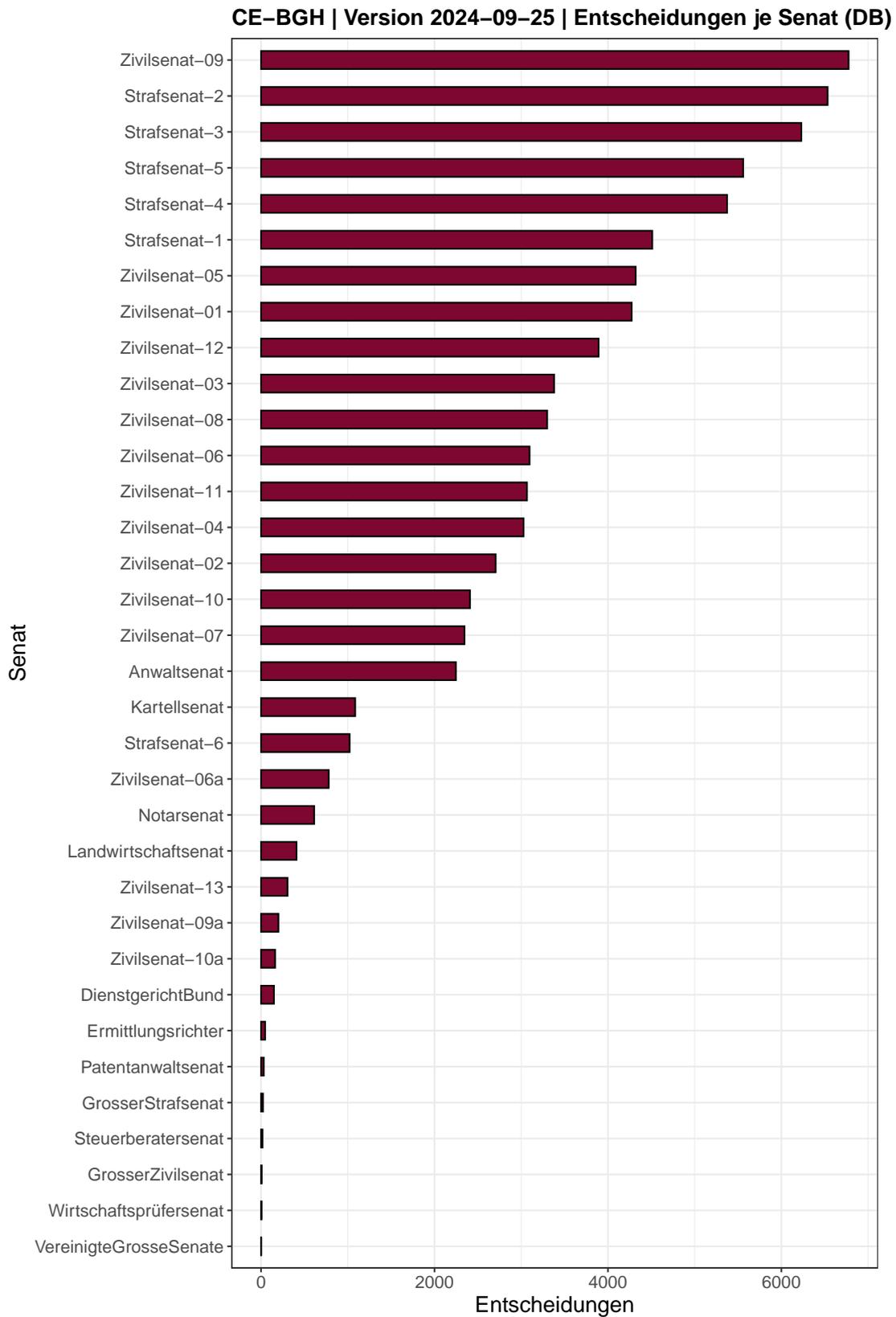


Senat	Entscheidungen	% Gesamt	% Kumulativ
NA	5506	7.07	7.07
1	4530	5.82	12.88
2	6545	8.40	21.29
3	5315	6.82	28.11
4	5375	6.90	35.01
5	5560	7.14	42.15
6	1025	1.32	43.47
I	4275	5.49	48.95
II	2706	3.47	52.43
III	3377	4.34	56.76
IV	3027	3.89	60.65
IX	6775	8.70	69.35
IXa	201	0.26	69.61
V	4320	5.55	75.15
VI	3095	3.97	79.12
VII	2347	3.01	82.14
VIII	3299	4.24	86.37
VIa	780	1.00	87.37
X	2409	3.09	90.47

(continued)

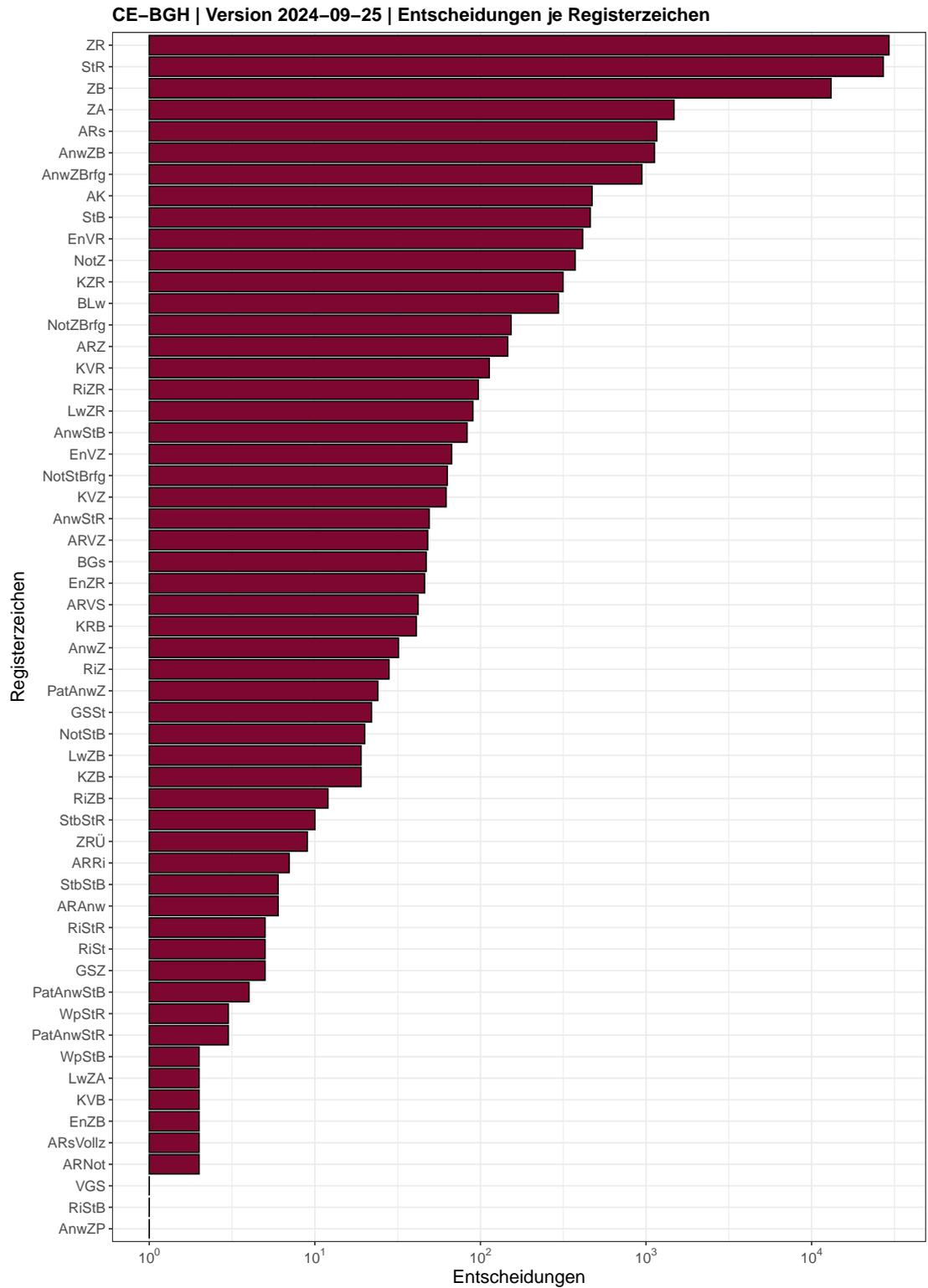
Senat	Entscheidungen	% Gesamt	% Kumulativ
XI	3066	3.94	94.40
XII	3891	5.00	99.40
XIII	306	0.39	99.79
Xa	162	0.21	100.00
Total	77892	100.00	100.00

11.4 Nach Spruchkörper (Datenbank)



Senat	Entscheidungen	% Gesamt	% Kumulativ
Anwaltsenat	2247	2.88	2.88
DienstgerichtBund	148	0.19	3.07
Ermittlungsrichter	47	0.06	3.14
GrosserStrafsenat	22	0.03	3.16
GrosserZivilsenat	6	0.01	3.17
Kartellsenat	1085	1.39	4.56
Landwirtschaftsenat	409	0.53	5.09
Notarsenat	612	0.79	5.87
Patentanwaltsenat	31	0.04	5.91
Steuerberatersenat	16	0.02	5.94
Strafsenat-1	4511	5.79	11.73
Strafsenat-2	6533	8.39	20.11
Strafsenat-3	6230	8.00	28.11
Strafsenat-4	5374	6.90	35.01
Strafsenat-5	5559	7.14	42.15
Strafsenat-6	1021	1.31	43.46
VereinigteGrosseSenate	1	0.00	43.46
Wirtschaftsprüfersenat	5	0.01	43.47
Zivilsenat-01	4274	5.49	48.95
Zivilsenat-02	2705	3.47	52.43
Zivilsenat-03	3379	4.34	56.76
Zivilsenat-04	3027	3.89	60.65
Zivilsenat-05	4320	5.55	66.20
Zivilsenat-06	3096	3.97	70.17
Zivilsenat-06a	780	1.00	71.17
Zivilsenat-07	2346	3.01	74.18
Zivilsenat-08	3299	4.24	78.42
Zivilsenat-09	6775	8.70	87.12
Zivilsenat-09a	200	0.26	87.37
Zivilsenat-10	2409	3.09	90.47
Zivilsenat-10a	162	0.21	90.68
Zivilsenat-11	3066	3.94	94.61
Zivilsenat-12	3892	5.00	99.61
Zivilsenat-13	305	0.39	100.00
Total	77892	100.00	100.00

11.5 Nach Registerzeichen

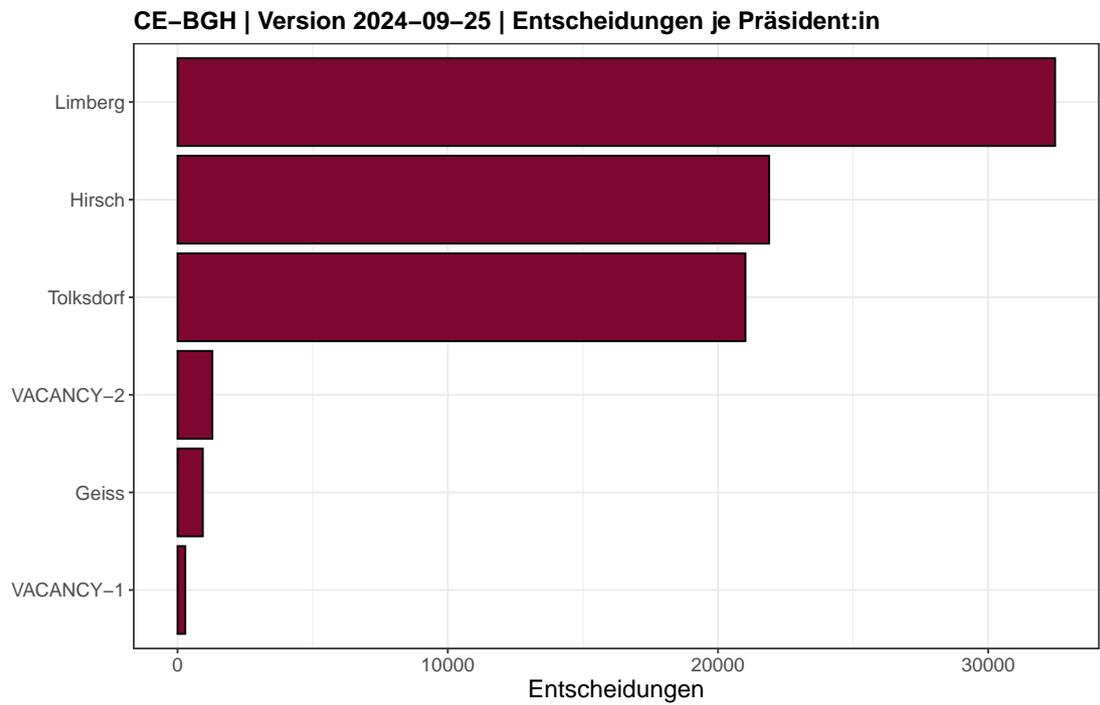


Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
AK	472	0.61	0.61
ARAnw	6	0.01	0.61
ARNot	2	0.00	0.62
ARRi	7	0.01	0.63
ARVS	42	0.05	0.68
ARVZ	48	0.06	0.74
ARZ	146	0.19	0.93
ARs	1161	1.49	2.42
ARsVollz	2	0.00	2.42
AnwStB	83	0.11	2.53
AnwStR	49	0.06	2.59
AnwZ	32	0.04	2.63
AnwZB	1125	1.44	4.08
AnwZBrfg	944	1.21	5.29
AnwZP	1	0.00	5.29
BGs	47	0.06	5.35
BLw	296	0.38	5.73
EnVR	414	0.53	6.26
EnVZ	67	0.09	6.35
EnZB	2	0.00	6.35
EnZR	46	0.06	6.41
GSSSt	22	0.03	6.44
GSZ	5	0.01	6.44
KRB	41	0.05	6.50
KVB	2	0.00	6.50
KVR	113	0.15	6.64
KVZ	62	0.08	6.72
KZB	19	0.02	6.75
KZR	315	0.40	7.15
LwZA	2	0.00	7.15
LwZB	19	0.02	7.18
LwZR	90	0.12	7.29
NotStB	20	0.03	7.32
NotStBrfg	63	0.08	7.40
NotZ	373	0.48	7.88
NotZBrfg	153	0.20	8.08
PatAnwStB	4	0.01	8.08
PatAnwStR	3	0.00	8.09
PatAnwZ	24	0.03	8.12
RiSt	5	0.01	8.12

(continued)

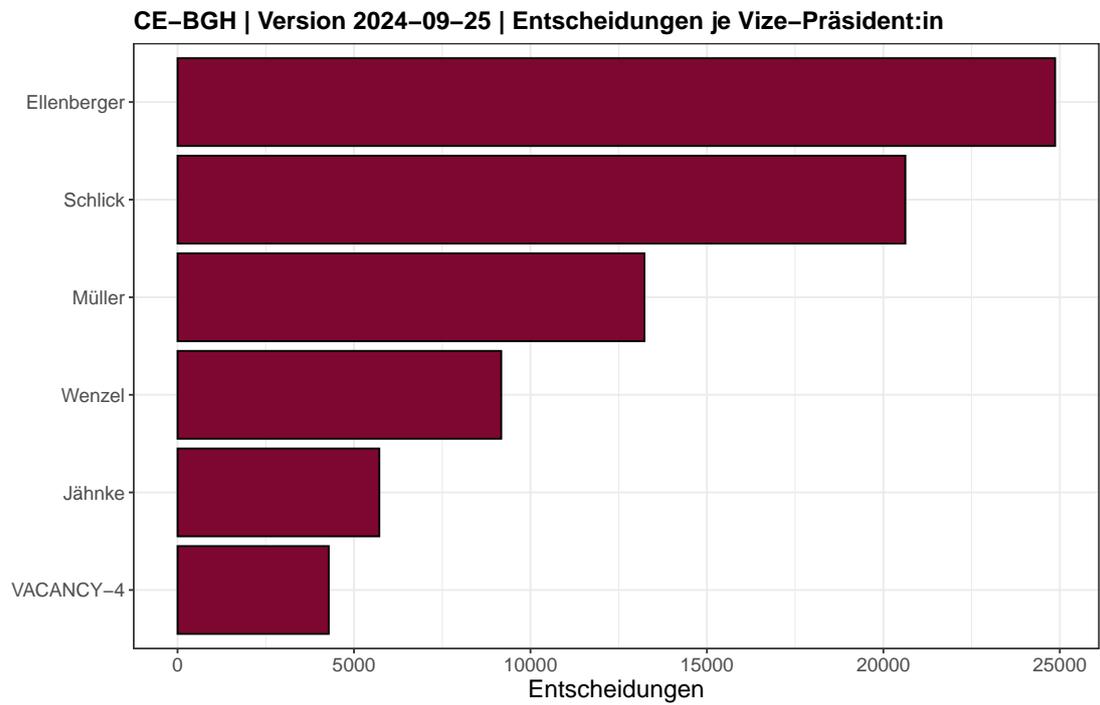
Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
RiStB	1	0.00	8.12
RiStR	5	0.01	8.13
RiZ	28	0.04	8.17
RiZB	12	0.02	8.18
RiZR	97	0.12	8.31
StB	460	0.59	8.90
StR	27076	34.76	43.66
StbStB	6	0.01	43.67
StbStR	10	0.01	43.68
VGS	1	0.00	43.68
WpStB	2	0.00	43.68
WpStR	3	0.00	43.69
ZA	1475	1.89	45.58
ZB	13096	16.81	62.39
ZR	29284	37.60	99.99
ZRÜ	9	0.01	100.00
Total	77892	100.00	100.00

11.6 Nach Präsident:in



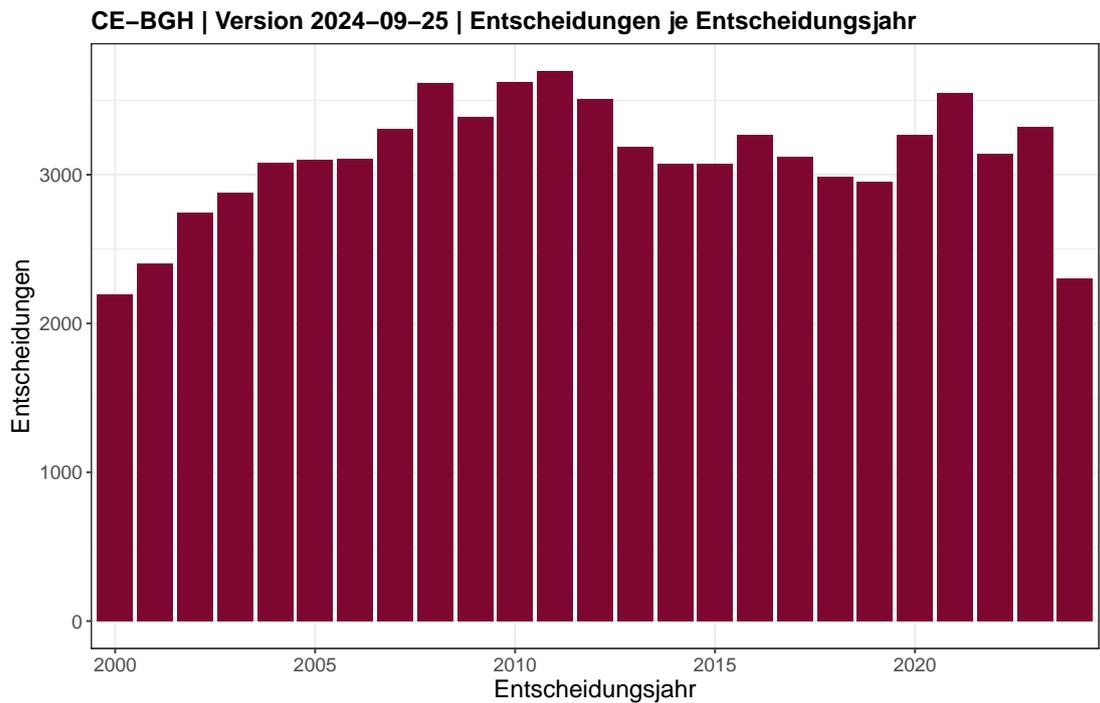
Präsident:in	Entscheidungen	% Gesamt	% Kumulativ
Geiss	935	1.20	1.20
Hirsch	21893	28.11	29.31
Limberg	32473	41.69	71.00
Tolksdorf	21014	26.98	97.98
VACANCY-1	288	0.37	98.35
VACANCY-2	1289	1.65	100.00
Total	77892	100.00	100.00

11.7 Nach Vize-Präsident:in



Vize-Präsident:in	Entscheidungen	% Gesamt	% Kumulativ
Ellenberger	24866	31.92	31.92
Jähnke	5717	7.34	39.26
Müller	13230	16.99	56.25
Schlick	20623	26.48	82.72
VACANCY-4	4284	5.50	88.22
Wenzel	9172	11.78	100.00
Total	77892	100.00	100.00

11.8 Nach Entscheidungsjahr

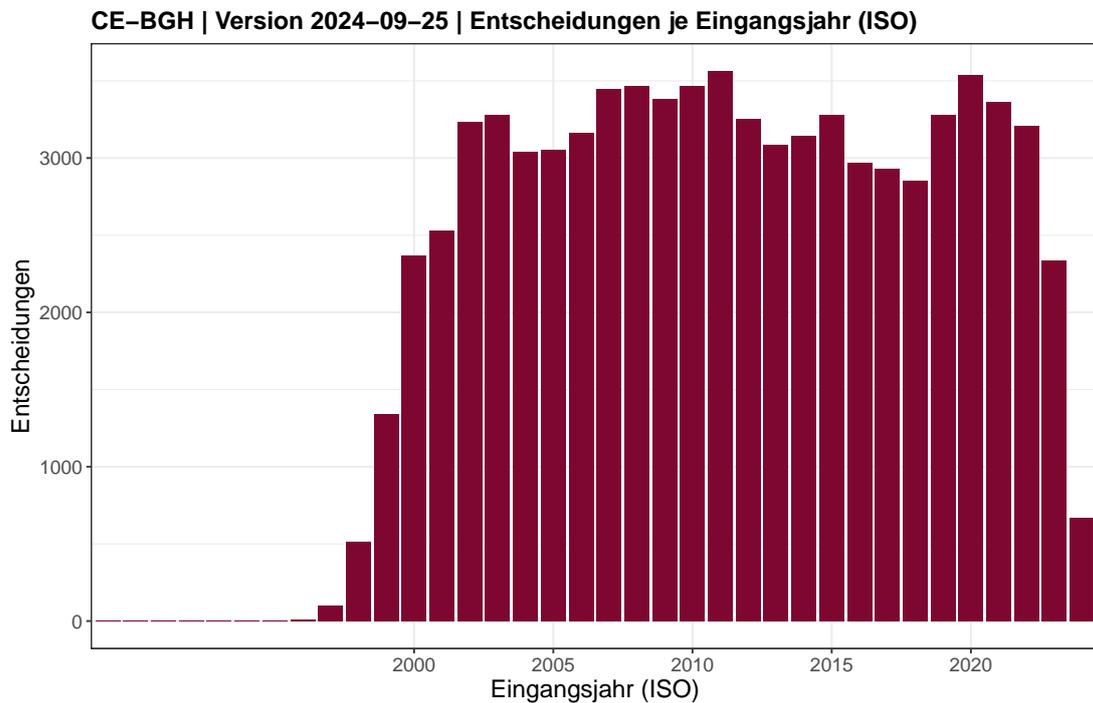


Entscheidungsjahr	Entscheidungen	% Gesamt	% Kumulativ
2000	2197	2.82	2.82
2001	2404	3.09	5.91
2002	2746	3.53	9.43
2003	2881	3.70	13.13
2004	3079	3.95	17.08
2005	3098	3.98	21.06
2006	3106	3.99	25.05
2007	3310	4.25	29.30
2008	3614	4.64	33.94
2009	3391	4.35	38.29
2010	3623	4.65	42.94
2011	3697	4.75	47.69
2012	3506	4.50	52.19
2013	3190	4.10	56.29
2014	3071	3.94	60.23
2015	3074	3.95	64.17
2016	3271	4.20	68.37
2017	3120	4.01	72.38
2018	2988	3.84	76.22

(continued)

Entscheidungsjahr	Entscheidungen	% Gesamt	% Kumulativ
2019	2951	3.79	80.00
2020	3265	4.19	84.20
2021	3547	4.55	88.75
2022	3141	4.03	92.78
2023	3322	4.26	97.05
2024	2300	2.95	100.00
Total	77892	100.00	100.00

11.9 Nach Eingangsjahr (ISO)



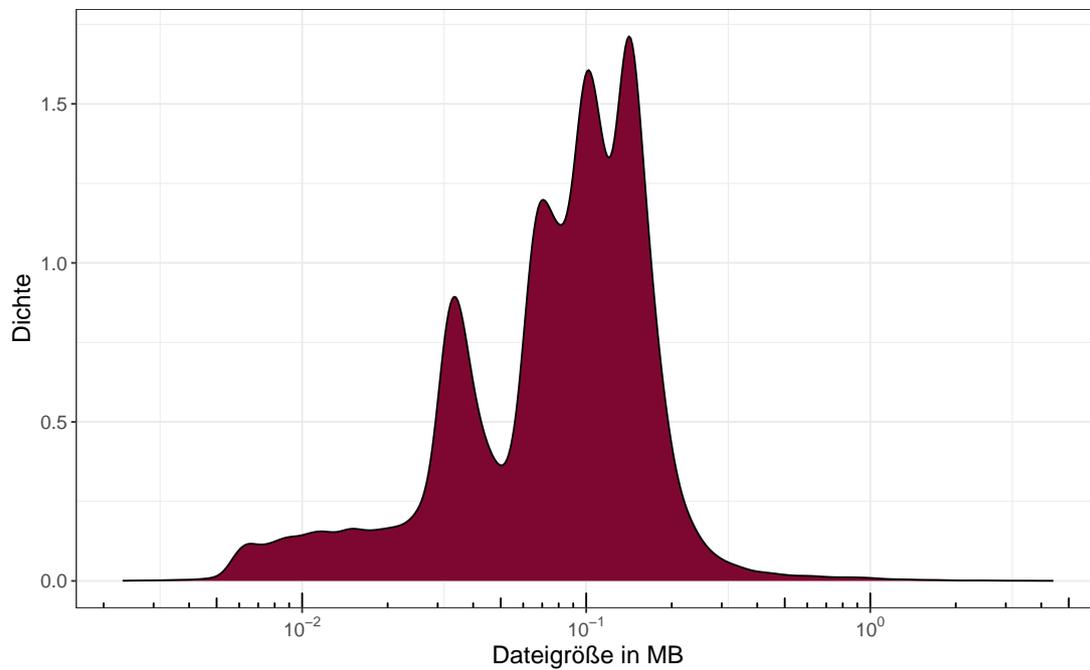
Eingangsjahr	Entscheidungen	% Gesamt	% Kumulativ
1980	1	0.00	0.00
1986	1	0.00	0.00
1988	1	0.00	0.00
1991	1	0.00	0.01
1993	2	0.00	0.01
1994	2	0.00	0.01
1995	4	0.01	0.02
1996	9	0.01	0.03
1997	100	0.13	0.16
1998	515	0.66	0.82
1999	1343	1.72	2.54
2000	2368	3.04	5.58
2001	2530	3.25	8.83
2002	3234	4.15	12.98
2003	3280	4.21	17.19
2004	3038	3.90	21.09
2005	3055	3.92	25.01
2006	3161	4.06	29.07
2007	3447	4.43	33.50

(continued)

Eingangsjahr	Entscheidungen	% Gesamt	% Kumulativ
2008	3465	4.45	37.95
2009	3384	4.34	42.29
2010	3466	4.45	46.74
2011	3564	4.58	51.32
2012	3257	4.18	55.50
2013	3086	3.96	59.46
2014	3144	4.04	63.50
2015	3280	4.21	67.71
2016	2971	3.81	71.52
2017	2932	3.76	75.29
2018	2853	3.66	78.95
2019	3282	4.21	83.16
2020	3538	4.54	87.70
2021	3365	4.32	92.02
2022	3207	4.12	96.14
2023	2334	3.00	99.14
2024	672	0.86	100.00
Total	77892	100.00	100.00

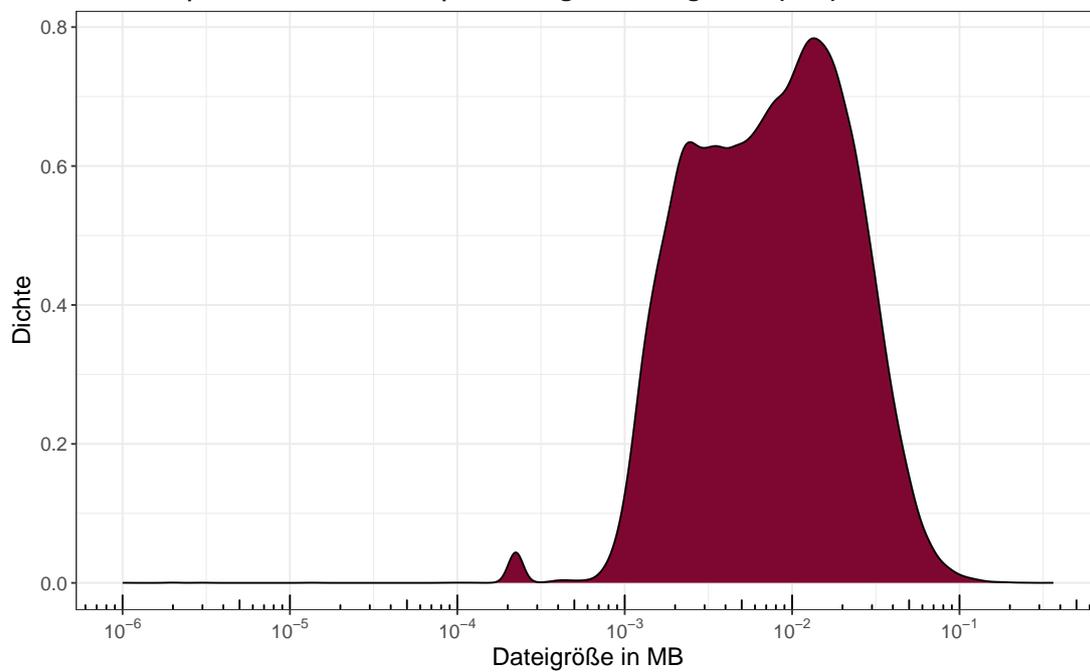
12 Dateigrößen

CE-BGH | Version 2024-09-25 | Verteilung der Dateigrößen (PDF)



Fobbe | DOI: 10.5281/zenodo.12814022

CE-BGH | Version 2024-09-25 | Verteilung der Dateigrößen (TXT)



Fobbe | DOI: 10.5281/zenodo.12814022

13 Kryptographische Signaturen

13.1 Zwei-Phasen-Signatur

Die Integrität und Echtheit der einzelnen Archive des Datensatzes sind durch eine Zwei-Phasen-Signatur sichergestellt.

In **Phase I** werden während der Kompilierung für jedes ZIP-Archiv, das Codebook und die Robustness Checks Hash-Werte in zwei verschiedenen Verfahren (SHA2-256 und SHA3-512) berechnet und in einer CSV-Datei dokumentiert.

In **Phase II** werden diese CSV-Datei und der Compilation Report mit meinem persönlichen geheimen GPG-Schlüssel signiert. Dieses Verfahren stellt sicher, dass die Kompilierung von jedermann durchgeführt werden kann, insbesondere im Rahmen von Replikationen, die persönliche Gewähr für Ergebnisse aber dennoch vorhanden bleibt.

13.2 Persönliche GPG-Signatur

Die während der Kompilierung des Datensatzes erstellte CSV-Datei mit den Hash-Prüfsummen und der Compilation Report sind mit meiner persönlichen GPG-Signatur versehen. Der mit dieser Version korrespondierende Public Key ist sowohl mit dem Datensatz als auch mit dem Source Code hinterlegt. Er hat folgende Kenndaten:

Name: Sean Fobbe (fobbe-data@posteo.de)

Fingerabdruck: FE6F B888 F0E5 656C 1D25 3B9A 50C4 1384 F44A 4E42

14 Changelog

14.1 Version 2024-09-25

- Vollständige Aktualisierung der Daten
- LIZENZÄNDERUNG: Source Code jetzt unter GNU General Public License Version 3 (GPLv3) oder später lizenziert
- NEU: Zitationsnetzwerk Aktenzeichen-zu-BGHZ und Aktenzeichen zu BGHSt
- NEU: Variable für BGHZ
- Extraktion von Zitaten für Aktenzeichen-zu-Aktenzeichen komplett überarbeitet
- R-Version auf 4.4.0 aktualisiert (wegen CVE-2024-27322)
- Variable “bghr” wird nun korrekt extrahiert
- Source-Archiv wird nun aus dem Git-Manifest gebildet
- Anpassung von Compose File an Debian 11
- Docker Zeitzone auf Berlin eingestellt
- Vereinfachung der Repository-Struktur
- Python Toolchain entfernt
- Zusätzliches Lösch-Skript mit Docker-Integration
- Verbesserung von Warnmeldungen
- Erweiterung der Unit Tests
- Erweiterung der automatischen Korrektur der Aktenzeichen
- Extraktion von PDFs ignoriert fehlerhafte PDFs
- Fix für Segmentation Fault bei Zählung von Types bei leeren Dokumenten
- Fix für gitconfig-Problem
- Diagramme nicht mehr nummeriert, sondern nach Typ geordnet

14.2 Version 2023-03-10

- Vollständige Aktualisierung der Daten
- NEU Zitations-Netzwerk zwischen allen Aktenzeichen des Bundesgerichtshofs als GraphML zur freien Verwendung (EXPERIMENTELL!)
- Gesamte Laufzeitumgebung mit Docker versionskontrolliert
- Aktenzeichen aus dem Eingangszeitraum 2000 bis 2009 nun korrekt mit führender Null formatiert (z.B. 1 BvR 44/02 statt 1 BvR 44/2)
- Vereinfachung der Konfigurations-Datei
- Verbesserte Formatierung von Warnungen und Fehlermeldungen im Compilation Report
- Verbesserung des Robustness Check Reports
- Neue Funktion für automatischen clean run (Löschung aller Zwischenergebnisse)
- Update der Download-Funktion
- Überflüssige Warnung in f.future_lingsummarize-Funktion entfernt
- Alle Roh-Dateien werden nun im Unterordner “files” gespeichert
- Korrektur für RiSt-Aktenzeichen eingefügt

14.3 Version 2022-08-16

- Vollständige Aktualisierung der Daten
- Neuer Entwurf des gesamten Source Codes im {targets} framework

- Die Zivilsenate in der Variable “spruchkoerper_db” sind jetzt arabisch nummeriert, damit sie automatisch sortiert werden können. Die originale römische Nummerierung ist weiterhin in der Variable “spruchkoerper_az” zu finden.
- Variable “comment” wurde in “bemerkung” umbenannt
- Variable “berichtigung” ist jetzt entweder mit TRUE oder FALSE codiert
- Diagramme sind in neuer Reihenfolge nummeriert, um die Reihenfolge im Codebook abzubilden
- Umfang der Datenbankabfrage ist nun vollautomatisiert

14.4 Version 2022-02-12

- Vollständige Aktualisierung der Daten
- Strenge Kontrolle und semantische Sortierung aller Variablen (entsprechend der Reihenfolge im Codebook)
- Datenstruktur wird nicht mehr im Codebook angezeigt um Fehler mit der UTF8-Kodierung und *listings* für L^AT_EX zu vermeiden
- Strenge Versionskontrolle aller R packages mit *renv*
- Der Prozess der Kompilierung ist jetzt detailliert konfigurierbar, insbesondere die Parallelisierung
- Parallelisierung nun vollständig mit *future* statt mit *foreach* und *doParallel*
- Fehlerhafte Kompilierungen werden beim vor der nächsten Kompilierung vollautomatisch aufgeräumt
- Alle Ergebnisse werden automatisch fertig verpackt in den Ordner ‘output’ sortiert
- README und CHANGELOG sind jetzt externe Markdown-Dateien, die bei der Kompilierung automatisch eingebunden werden
- Issue #1 fixed: Senate normalisiert; die Variable “spruchkoerper_db” enthält nun die Präfixe “Strafsenat” und “Zivilsenat” vor der jeweiligen Senatsnummer um in den Dateinamen eine einfachere Orientierung zu ermöglichen
- Issue #2 fixed: Variablen nicht mehr doppelt definiert
- Issue #3 fixed: Alle Dateinamen-Präfixe nun korrekt
- Source Code des Changelogs zu Markdown konvertiert
- In der Vergangenheit fälschlich als “Platzhalter” aussortierte drei Dokumente sind nun im Datensatz enthalten
- Das Diagramm “Entscheidungen je Registerzeichen” ist nun zu einer Log-Skala konvertiert um die Darstellung informativer zu gestalten

14.5 Version 2021-04-27

- Vollständige Aktualisierung der Daten
- Veröffentlichung des vollständigen Source Codes
- Deutliche Erweiterung des inhaltlichen Umfangs des Codebooks
- Einführung der vollautomatischen Erstellung von Datensatz und Codebook
- Einführung von Compilation Reports um den Erstellungsprozess exakt zu dokumentieren
- Einführung von Variablen für Lizenz, Versionsnummer, Concept DOI, Version DOI, ECLI, Typ der Entscheidung, Präsident:in, Vize-Präsident:in, Verfahrensart, Name, Leitsatz, Bemerkungen, Berichtigungen, und linguistische Kennzahlen (Zeichen, Tokens, Typen, Sätze)
- Einführung von PDF-Varianten für Leitsatzentscheidungen und namentlich gekennzeichneten Entscheidungen.

- Zusammenfügung von über Zeilengrenzen getrennten Wörtern in der CSV-Variante
- Automatisierung und Erweiterung der Qualitätskontrolle
- Einführung von Diagrammen zur Visualisierung von Prüfergebnissen
- Einführung kryptographischer Signaturen
- Alle Variablen sind nun in Kleinschreibung und Snake Case gehalten
- Variable ‘Ordinalzahl’ in ‘eingangsnummer’ umbenannt

14.6 Version 2020-07-09

- Erstveröffentlichung

15 Parameter für strenge Replikationen

```
## [1] "OpenSSL 3.0.2 15 Mar 2022 (Library: OpenSSL 3.0.2 15 Mar 2022)"
```

```
## R version 4.4.0 (2024-04-24)
## Platform: x86_64-pc-linux-gnu
## Running under: Ubuntu 22.04.4 LTS
##
## Matrix products: default
## BLAS: /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblas-p-r0.3.20.so;
## LAPACK version 3.10.0
##
## locale:
## [1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C
## [3] LC_TIME=en_US.UTF-8 LC_COLLATE=en_US.UTF-8
## [5] LC_MONETARY=en_US.UTF-8 LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=en_US.UTF-8 LC_NAME=C
## [9] LC_ADDRESS=C LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Europe/Berlin
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats graphics grDevices utils datasets methods base
##
## other attached packages:
## [1] future.apply_1.11.2 future_1.34.0 quanteda_4.1.0
## [4] readtext_0.91 data.table_1.16.0 scales_1.3.0
## [7] ggraph_2.2.1 igraph_2.0.3 ggplot2_3.5.1
## [10] pdftools_3.4.1 kableExtra_1.4.0 knitr_1.48
## [13] rvest_1.0.4 httr_1.4.7 mgsub_1.7.3
## [16] zip_2.3.1 fs_1.6.4 testthat_3.2.1.1
## [19] RcppTOML_0.2.2 tarchetypes_0.9.0 targets_1.7.1
##
## loaded via a namespace (and not attached):
## [1] tidyselect_1.2.1 viridisLite_0.4.2 dplyr_1.1.4
## [4] farver_2.1.2 viridis_0.6.5 fastmap_1.2.0
## [7] tweenr_2.0.3 stringfish_0.16.0 digest_0.6.37
## [10] base64url_1.4 lifecycle_1.0.4 secretbase_1.0.2
## [13] qpdf_1.3.3 waldo_0.5.3 processx_3.8.4
## [16] magrittr_2.0.3 compiler_4.4.0 rlang_1.1.4
## [19] tools_4.4.0 utf8_1.2.4 yaml_2.3.10
## [22] labeling_0.4.3 askpass_1.2.0 stopwords_2.3
## [25] graphlayouts_1.1.1 curl_5.2.3 xml2_1.3.6
## [28] pkgload_1.4.0 withr_3.0.1 purrr_1.0.2
## [31] desc_1.4.3 grid_4.4.0 polyclip_1.10-7
## [34] fansi_1.0.6 colorspace_2.1-1 globals_0.16.3
## [37] MASS_7.3-60.2 tinytex_0.53 cli_3.6.3
## [40] rmarkdown_2.28 generics_0.1.3 RcppParallel_5.1.9
## [43] rstudioapi_0.16.0 RApiSerialize_0.1.3 cachem_1.1.0
## [46] ggforce_0.4.2 stringr_1.5.1 parallel_4.4.0
```

```
## [49] selectr_0.4-2      vctrs_0.6.5      Matrix_1.7-0
## [52] callr_3.7.6        ggrepel_0.9.6    listenv_0.9.1
## [55] systemfonts_1.1.0  tidyr_1.3.1      glue_1.7.0
## [58] parallely_1.38.0   codetools_0.2-20 ps_1.8.0
## [61] stringi_1.8.4      gtable_0.3.5     munsell_0.5.1
## [64] tibble_3.2.1       pillar_1.9.0     htmltools_0.5.8.1
## [67] brio_1.1.5         R6_2.5.1         rprojroot_2.0.4
## [70] tidygraph_1.3.1    evaluate_1.0.0   lattice_0.22-6
## [73] qs_0.26.3          backports_1.5.0  memoise_2.0.1
## [76] Rcpp_1.0.13        fastmatch_1.1-4  svglite_2.1.3
## [79] gridExtra_2.3      xfun_0.47        pkgconfig_2.0.3
```

Literaturverzeichnis

- Allaire, JJ, Yihui Xie, Christophe Dervieux, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, et al. 2024. *Rmarkdown: Dynamic Documents for R*. <https://github.com/rstudio/rmarkdown>.
- Barrett, Tyson, Matt Dowle, Arun Srinivasan, Jan Gorecki, Michael Chirico, Toby Hocking, and Benjamin Schwendinger. 2024. *Data.table: Extension of 'Data.frame'*. <https://r-datatable.com>.
- Bengtsson, Henrik. 2021. "A Unifying Framework for Parallel and Distributed Processing in R Using Futures." *The R Journal* 13 (2): 208–27. <https://doi.org/10.32614/RJ-2021-048>.
- . 2024a. *Future.apply: Apply Function to Elements in Parallel Using Futures*. <https://future.apply.futureverse.org>.
- . 2024b. *Future: Unified Parallel and Distributed Processing in R for Everyone*. <https://future.futureverse.org>.
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. "Quanteda: An R Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3 (30): 774. <https://doi.org/10.21105/joss.00774>.
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, Akitaka Matsuo, and William Lowe. 2024. *Quanteda: Quantitative Analysis of Textual Data*. <https://quanteda.io>.
- Csárdi, Gábor. 2024. *Zip: Cross-Platform Zip Compression*. <https://github.com/r-lib/zip>.
- Csardi, Gabor, and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal Complex Systems*: 1695. <https://igraph.org>.
- Csárdi, Gábor, Tamás Nepusz, Vincent Traag, Szabolcs Horvát, Fabio Zanini, Daniel Noom, and Kirill Müller. 2024. *Igraph: Network Analysis and Visualization*. <https://r.igraph.org/>.
- Eddelbuettel, Dirk. 2023. *RcppTOML: Rcpp Bindings to Parser for "Tom's Obvious Markup Language"*. <http://dirk.eddelbuettel.com/code/rcpp.toml.html>.
- Ewing, Mark. 2021. *Mgsub: Safe, Multiple, Simultaneous String Substitution*.
- Gagolewski, Marek. 2022. "stringi: Fast and Portable Character String Processing in R." *Journal of Statistical Software* 103 (2): 1–59. <https://doi.org/10.18637/jss.v103.i02>.
- Gagolewski, Marek, Bartek Tartanus, others; Unicode, Inc., and others. 2024. *Stringi: Fast and Portable Character String Processing Facilities*. <https://stringi.gagolewski.com/>.
- Landau, William Michael. 2021a. *Tarchetypes: Archetypes for Targets*.
- . 2021b. "The Targets R Package: A Dynamic Make-Like Function-Oriented Pipeline Toolkit for Reproducibility and High-Performance Computing." *Journal of Open Source Software* 6 (57): 2959. <https://doi.org/10.21105/joss.02959>.
- . 2024a. *Tarchetypes: Archetypes for Targets*. <https://docs.ropensci.org/tarchetypes/>.
- . 2024b. *Targets: Dynamic Function-Oriented Make-Like Declarative Pipelines*. <https://docs.ropensci.org/targets/>.

- Ooms, Jeroen. 2024. *Pdftools: Text Extraction, Rendering and Converting of Pdf Documents*. <https://docs.ropensci.org/pdftools/>.
- Pedersen, Thomas Lin. 2024. *Ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*. <https://ggraph.data-imaginist.com>.
- Ushey, Kevin, and Hadley Wickham. 2024. *Renv: Project Environments*. <https://rstudio.github.io/renv/>.
- Wickham, Hadley. 2024. *Rvest: Easily Harvest (Scrape) Web Pages*. <https://rvest.tidyverse.org/>.
- Xie, Yihui. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.
- . 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.
- . 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://yihui.org/knitr/>.
- Xie, Yihui, J. J. Allaire, and Garrett Golemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown>.
- Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown-cookbook>.
- Zhu, Hao. 2024. *KableExtra: Construct Complex Table with Kable and Pipe Syntax*. <http://haozhu233.github.io/kableExtra/>.