

**Corpus der Entscheidungen  
des  
Bundesverfassungsgerichts  
(CE-BVerfG)**

CODEBOOK

Version 2021-01-08



DOI: [10.5281/zenodo.4308215](https://doi.org/10.5281/zenodo.4308215)

<b>Titel</b>	Corpus der Entscheidungen des Bundesverfassungsgerichts
<b>Abkürzung</b>	CE-BVerfG
<b>Autor</b>	Seán Fobbe
<b>Version</b>	2021-01-08
<b>Download</b>	<a href="https://doi.org/10.5281/zenodo.4308215">https://doi.org/10.5281/zenodo.4308215</a>
<b>Lizenz</b>	CC0 1.0 Universal

### Zitiervorschlag

*Seán Fobbe* (2021). Corpus der Entscheidungen des Bundesverfassungsgerichts (CE-BVerfG). Version 2021-01-08. Zenodo. DOI: 10.5281/zenodo.4308215.

### 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 2021-01-08. 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.3902658. 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

<b>1</b>	<b>Einführung</b>	<b>5</b>
<b>2</b>	<b>Nutzung</b>	<b>6</b>
2.1	CSV-Dateien . . . . .	6
2.2	TXT-Dateien . . . . .	6
<b>3</b>	<b>Konstruktion</b>	<b>7</b>
3.1	Datenquellen . . . . .	7
3.2	Beschreibung des Datensatzes . . . . .	7
3.3	Sammlung der Daten . . . . .	7
3.4	Source Code und Compilation Report . . . . .	7
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 . . . . .	8
3.7.3	Beispiel eines Dateinamens . . . . .	9
3.8	Qualitätsprüfung . . . . .	9
3.9	Grafische Darstellung . . . . .	9
<b>4</b>	<b>Varianten und Zielgruppen</b>	<b>10</b>
<b>5</b>	<b>Variablen</b>	<b>12</b>
5.1	Hinweise . . . . .	12
5.2	Erläuterungen der einzelnen Variablen . . . . .	12
<b>6</b>	<b>Registerzeichen</b>	<b>17</b>
<b>7</b>	<b>PräsidentInnen</b>	<b>18</b>
7.1	Hinweise . . . . .	18
7.2	Lebensdaten . . . . .	18
7.3	Dienstalter und Lebensalter . . . . .	18
<b>8</b>	<b>Vize-PräsidentInnen</b>	<b>19</b>
8.1	Hinweise . . . . .	19
8.2	Lebensdaten . . . . .	19
8.3	Dienstalter und Lebensalter . . . . .	20
<b>9</b>	<b>Linguistische Kennzahlen</b>	<b>21</b>
9.1	Erläuterung der Kennzahlen und Diagramme . . . . .	21
9.2	Werte der Kennzahlen . . . . .	21
9.3	Verteilung Tokens . . . . .	22
9.4	Verteilung Typen . . . . .	22
9.5	Verteilung Sätze . . . . .	23
<b>10</b>	<b>Inhalt des Korpus</b>	<b>24</b>
10.1	Zusammenfassung . . . . .	24
10.2	Nach Spruchkörper (Aktenzeichen) . . . . .	24
10.3	Nach Registerzeichen . . . . .	25

10.4 Nach PräsidentIn . . . . .	27
10.5 Nach Vize-PräsidentIn . . . . .	28
10.6 Nach Entscheidungsjahr . . . . .	29
10.7 Nach Eingangsjahr (ISO) . . . . .	31
<b>11 Dateigrößen</b>	<b>33</b>
<b>12 Signaturprüfung</b>	<b>34</b>
12.1 Allgemeines . . . . .	34
12.2 Persönliche GPG-Signatur . . . . .	34
12.3 Import: Public Key . . . . .	34
12.4 Prüfung: GPG-Signatur der Hash-Datei . . . . .	35
12.5 Prüfung: SHA3-512 Hashes der ZIP-Archive . . . . .	36
<b>13 Changelog</b>	<b>37</b>
<b>14 Parameter für strenge Replikationen</b>	<b>38</b>
<b>Literaturverzeichnis</b>	<b>39</b>

# 1 Einführung

Das **Bundesverfassungsgericht (BVerfG)** ist das höchste Gericht der Bundesrepublik Deutschland und ein Verfassungsorgan. Als »Hüter der Verfassung« ist es seit seiner Gründung im Jahr 1951 mit der Auslegung und Durchsetzung des Grundgesetzes betraut.

Seine Bedeutung im Verfassungsgefüge der Bundesrepublik Deutschland ist kaum zu überschätzen. So richtet es nicht nur über Streitigkeiten zwischen Verfassungsorganen und über Normenkontrollanträge, welche die Nichtigkeit von Gesetzen zur Folge haben können, sondern auch über »Verfassungsbeschwerden, die von jedermann mit der Behauptung erhoben werden können, durch die öffentliche Gewalt in einem seiner Grundrechte oder in einem seiner [grundrechtsgleichen Rechte] verletzt zu sein« (Art. 93 Abs. 4b GG). Das Instrument der Verfassungsbeschwerde ist in seiner Beliebtheit und Effektivität in der Geschichte Deutschlands beispiellos und von hoher wissenschaftlicher und praktischer Bedeutung. In nicht wenigen Verfahrensarten haben die Entscheidungen des BVerfG zudem Gesetzeskraft (§ 31 Abs. 2 BVerfGG).

Die quantitative Analyse von juristischen Texten, insbesondere denen des Bundesverfassungsgerichts, ist in den deutschen Rechtswissenschaften ein noch junges und kaum bearbeitetes Feld.<sup>1</sup> 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*<sup>2</sup> und *openJur*<sup>3</sup> verbieten ausdrücklich das maschinelle Auslesen der Rohdaten.<sup>4</sup> 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 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.<sup>5</sup>

---

<sup>1</sup> Besonders positive Ausnahmen finden sich unter: <https://www.quantitative-rechtswissenschaft.de/>

<sup>2</sup> <https://opiniojuris.de/>

<sup>3</sup> <https://openjur.de/>

<sup>4</sup> Openjur beabsichtigt eine API anzubieten, diese war aber im November 2020 immernoch nicht verfügbar. Openjur ist seit 2008 in Betrieb.

<sup>5</sup> 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. CSV-Dateien<sup>6</sup> sind 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.

Hier empfehle ich für **R** dringend 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 Vorschlag:

```
library(data.table)
dt.bverfg <- 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.bverfg <- readtext("*.txt",
  docvarsfrom = "filenames",
  docvarnames = c("gericht",
    "datum",
    "spruchkoerper_typ",
    "spruchkoerper_az",
    "registerzeichen",
    "eingangsnummer",
    "eingangsjahr_az",
    "kollision",
    "name",
    "band",
    "seite"),
  dvsep = "_",
  encoding = "UTF-8")
```

<sup>6</sup> Das CSV-Format ist in RFC 4180 definiert, siehe <https://tools.ietf.org/html/rfc4180>

## 3 Konstruktion

### 3.1 Datenquellen

Datenquelle	Fundstelle
Primäre Datenquelle	<a href="https://www.bundesverfassungsgericht.de">https://www.bundesverfassungsgericht.de</a>
Source Code	<a href="https://doi.org/10.5281/zenodo.4308217">https://doi.org/10.5281/zenodo.4308217</a>
Entscheidungsnamen	<a href="https://doi.org/10.5281/zenodo.4308217">https://doi.org/10.5281/zenodo.4308217</a>
BVerfGE-Fundstellen	<a href="https://doi.org/10.5281/zenodo.4308217">https://doi.org/10.5281/zenodo.4308217</a>
Lebensdaten	<a href="https://doi.org/10.5281/zenodo.4308217">https://doi.org/10.5281/zenodo.4308217</a>

**Hinweis:** Alle auf Zenodo bereitgestellten Datenquellen wurden vom Autor persönlich für die Zwecke dieses Datensatzes erstellt.

### 3.2 Beschreibung des Datensatzes

Dieser Datensatz ist eine digitale Zusammenstellung von möglichst allen Entscheidungen, die auf der amtlichen Internetpräsenz des Bundesverfassungsgerichts (BVerfGE) veröffentlicht sind. Er enthält alle Entscheidungen, die auf der offiziellen Webseite des Bundesverfassungsgerichts am jeweiligen Stichtag veröffentlicht waren. Die Stichtage für jede Version entsprechen exakt der Versionsnummer.

Zusätzlich zu den einfach maschinenlesbaren Formaten (TXT und CSV) sind die PDF-Rohdaten enthalten, damit Analysten 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.

### 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.4308217>

Mit jeder Kompilierung des vollständigen Datensatzes wird auch ein umfangreicher **Compilation Report** in einem attraktiv designten PDF-Format erstellt (ähnlich diesem Codebook). Der Compilation Report enthält den vollständigen Source Code, 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 des Erstellungs-Prozesses interessieren, lesen Sie diesen bitte zuerst.

### 3.5 Grenzen des Datensatzes

Nutzer sollten folgende wichtige Grenzen beachten:

1. Der Datensatz enthält nur das, was das Gericht auch tatsächlich veröffentlicht, nämlich begründete Entscheidungen (*publication bias*).
2. Es kann aufgrund technischer Grenzen bzw. Fehler sein, dass manche — im Grunde verfügbare — Entscheidungen nicht oder nicht korrekt abgerufen werden (*automation bias*).
3. Es werden nur PDF-Dateien abgerufen (*file type bias*). Manche Entscheidungen sind nur als HTML verfügbar. Die Metadaten der Entscheidungen ohne PDF-Datei werden explizit im Compilation Report dokumentiert.
4. Erst ab dem Jahr 1998 sind begründete Entscheidungen des BVerfG einigermaßen vollständig veröffentlicht, auch wenn frühere Entscheidungen vereinzelt auf der Webseite verfügbar sind (*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 wurden weitgehend aus den Hyperlinks zur jeweiligen Datei extrahiert. Hinzugefügt wurden von mir eine Reihe weitere Variablen, u.a. Gerichtsname, Entscheidungsname BVerfGE-Fundstelle, PräsidentIn, Vize-PräsidentIn, sowie Unter- und Trennstriche um die Maschinenlesbarkeit zu erleichtern. Der volle Satz an Metadaten ist nur in den CSV-Dateien enthalten. Alle hinzugefügten Metadaten sind zusammen mit dem Source Code vollständig maschinenlesbar dokumentiert und liegen entweder im CSV-Format vor oder sind direkt im Source Code enthalten.

Die Dateinamen der PDF- und TXT-Dateien enthalten Gerichtsname, Datum (Langform nach ISO-8601, d.h. YYYY-MM-DD), den Typ des Spruchkörpers, das offizielle Aktenzeichen, eine Kollisions-ID, den Namen der Entscheidung, sowie die BVerfGE-Fundstelle (Band und Seite).

#### 3.7.2 Schema für die Dateinamen

```
[gericht]_[datum]_[spruchkoerper_typ]_[spruchkoerper_az]_[registerzeichen]_[  
eingangsnummer]_[eingangsjahr_az]_[kollision]_[name]_[band]_[seite]
```



### 3.7.3 Beispiel eines Dateinamens

BVerfG\_1997-07-08\_S\_1\_BvR\_1243\_95\_NA\_Partielehrer\_96\_152.txt

## 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 Bundesverfassungsrichter ist »scharlachrot«. Der Hex-Wert hierfür ist #ca2129. Das ist besonders bei der Erstellung thematisch passender Graphen hilfreich. Alle im Compilation Report und diesem Codebook präsentierten Graphen sind in diesem scharlachrot gehalten.

## 4 Varianten und Zielgruppen

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

Lehrende werden zudem 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	<b>Traditionelle juristische Forschung.</b> Die PDF-Dokumente wie sie vom Bundesverfassungsgericht 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 unter Windows zu vermeiden und die Lesbarkeit zu verbessern.
CSV_Datensatz	<b>Legal Tech/Quantitative Forschung.</b> 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.
CSV_Metadaten	<b>Legal Tech/Quantitative Forschung.</b> Wie die andere CSV-Datei, nur ohne die Entscheidungstexte. Sinnvoll für AnalystInnen, die sich nur für die Metadaten interessieren und Speicherplatz sparen wollen.
TXT	<b>Subsidiär für alle Zielgruppen.</b> 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 qualitative ForscherInnen, die nur wenig Speicherplatz oder eine langsame Internetverbindung zur Verfügung haben oder für quantitative ForscherInnen, die beim Einlesen der CSV-Dateien Probleme haben.

---

Variante	Zielgruppe und Beschreibung
ANALYSE	<b>Alle Lehrenden und Forschenden.</b> 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. AnalystInnen mit fortgeschrittenen Kenntnissen in R können auch auf den Source Code zurückgreifen. Empfohlen für Nutzer die einzelne Inhalte aus dem Codebook für andere Zwecke (z.B. Präsentationen, eigene Publikationen) weiterverwenden möchten.

---

## 5 Variablen

### 5.1 Hinweise

- Fehlende Werte sind immer mit »NA« codiert
- Strings können grundsätzlich alle in UTF-8 definierten Zeichen (insbesondere Buchstaben, Zahlen und Sonderzeichen) enthalten.

### 5.2 Erläuterungen der einzelnen Variablen

Variable	Typ	Erläuterung
doc_id	String	(Nur CSV-Datei) Der Name der extrahierten TXT-Datei.
text	String	(Nur CSV-Datei) Der vollständige Inhalt der Entscheidung, so wie er in der von <a href="http://www.bundesverfassungsgericht.de">www.bundesverfassungsgericht.de</a> heruntergeladenen PDF-Datei dokumentiert ist.
gericht	Alphabetisch	In diesem Datensatz ist nur der Wert »BVerfG« vergeben. Dies ist der ECLI-Gerichtscode für »Bundesverfassungsgericht«. Diese Variable dient vor allem zur einfachen und transparenten Verbindung der Daten mit anderen Datensätzen.
datum	Datum	Das Datum der Entscheidung im Format YYYY-MM-DD (Langform nach ISO-8601). Die Langform ist für Menschen einfacher lesbar und wird maschinell auch öfter automatisch als Datumsformat erkannt.
spruchkoerper_typ	Alphabetisch	Der Typ des Spruchkörpers. Es sind die Werte »K« (Kammer), »S« (Senat), »P« (Plenum) und »B« (Beschwerdekammer gem. § 97c BVerfGG) vergeben.
spruchkoerper_az	Natürliche Zahl	Der im Aktenzeichen angegebene Spruchkörper. Es sind nur die Werte »1« und »2« vergeben. Die Werte stehen für den 1. oder 2. Senat des Gerichts. Für Verzögerungsentscheidungen der Beschwerdekammer ist der Wert »NA«. <b>Achtung:</b> Um die Entscheidungen eines bestimmten Senats zu analysieren reicht es nicht, die Variable »spruchkoerper_az« zu nutzen, es muss zusätzlich noch die Variable »spruchkoerper_typ« auf »S« gesetzt werden, weil ansonsten noch mit dem Senat assoziierte Entscheidungen seiner Kammern und des Plenums mit ausgewählt werden.

Variable	Typ	Erläuterung
registerzeichen	Alphabetisch	Das amtliche Registerzeichen. Es gibt die Verfahrensart an, in der die Entscheidung ergangen ist. Eine Erläuterung der Registerzeichen findet sich unter Punkt 6.
eingangsnummer	Natürliche Zahl	Verfahren des gleichen Eingangsjahres erhalten vom Gericht eine Nummer in der Reihenfolge ihres Eingangs. Die Zahl ist in den Dateinamen mit führenden Nullen (falls <1000) codiert.
eingangsjahr_az	Natürliche Zahl	Das im Aktenzeichen angegebene Jahr in dem das Verfahren beim Gericht anhängig wurde. Das Format ist eine zweistellige Jahreszahl (YY).
eingangsjahr_iso	Natürliche Zahl	(Nur CSV-Datei) Das nach ISO-8601 codierte Jahr in dem das Verfahren beim Bundesverfassungsgericht anhängig wurde. Das Format ist eine vierstellige Jahreszahl (YYYY), um eine maschinenlesbare und eindeutige Jahreszahl für den Eingang zur Verfügung zu stellen. Wurde aus der Variable »ingangsjahr_az« durch den Autor des Datensatzes berechnet, unter der Annahme, dass Jahreszahlen über 50 dem 20. Jahrhundert zugeordnet sind und andere Jahreszahlen dem 21. Jahrhundert.
entscheidungsjahr	Natürliche Zahl	(Nur CSV-Datei) Das Jahr in dem die Entscheidung ergangen ist. Das Format ist eine vierstellige Jahreszahl (YYYY). Wurde aus der Variable »datum« durch den Autor des Datensatzes berechnet.
kollision	Alphanumerisch	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 einem Kleinbuchstaben ausdifferenziert. Für die erste Entscheidung ist der Wert der Variable »NA«, also nicht vorhanden. Die zweite Entscheidung ist mit »a« identifiziert, die dritte mit »b« und so fort. In der offiziellen Beschreibung der ECLI-Ordinalzahl wird diese Variable als »Kollisionsnummer« bezeichnet. Buchstaben sind allerdings keine Nummern, daher die abweichende Bezeichnung. In einem einzigen Fall ist tatsächlich eine »1« vergeben. Es handelt sich hier vermutlich um einen Fehler.

Variable	Typ	Erläuterung
name	String	Der Name der Entscheidung. Für viele Entscheidungen aus der amtlichen Sammlung sind bekannte Namen vorhanden, diese wurden benutzt soweit möglich und auffindbar. Für weniger bekannte Entscheidungen wurde ein möglichst informativer Name vom Autor vergeben. Die konkrete Darstellung (ohne Leerzeichen, mit Bindestrichen usw.) ist Gründen der maschinellen Lesbarkeit geschuldet.
band	Natürliche Zahl	Der Band der amtlichen Sammlung in dem die Entscheidung veröffentlicht ist.
seite	Natürliche Zahl	Die genaue Fundstelle (Seitenzahl) der Entscheidung im jeweiligen Band der amtlichen Sammlung. Nur sinnvoll nutzbar im Zusammenspiel mit der Variable »band«.
praesi	Alphabetisch	(Nur CSV-Datei) Der Nachname des oder der PräsidentIn in dessen/deren Amtszeit das Datum der Entscheidung fällt.
v_praesi	Alphabetisch	(Nur CSV-Datei) Der Nachname des oder der Vize-PräsidentIn in dessen/deren Amtszeit das Datum der Entscheidung fällt.
aktenzeichen	String	(Nur CSV-Datei) Das amtliche Aktenzeichen. Die Variable wurde aus den Variablen »spruchkoerper_az«, »registerzeichen«, »eingangsnummer« und »eingangsjahr_az« durch den Autor des Datensatzes berechnet.
ecli	String	(Nur CSV-Datei) Der European Case Law Identifier (ECLI) der Entscheidung. Jeder Entscheidung ist eine einzigartige ECLI zugewiesen, ggf. mit Kollisions-ID. Die ECLI ist vor allem dann hilfreich, wenn dieser Datensatz mit anderen Datensätzen zusammengeführt werden und Doppelungen 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 »entscheidungsjahr«, »spruchkoerper_typ«, »datum«, »kollisions«, »spruchkoerper_az«, »registerzeichen«, »eingangsnummer« und »eingangsjahr_az« durch den Autor des Datensatzes berechnet.

Variable	Typ	Erläuterung
tokens	Natürliche Zahl	(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	Natürliche Zahl	(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.
saetze	Natürliche Zahl	(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	Datum	(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.

Variable	Typ	Erläuterung
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 <a href="http://www.doi.org">www.doi.org</a> immer die <b>aktuellste Version</b> 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 <b>konkreten Version</b> des Datensatzes. Dieser ist langzeit-stabil (persistent). Über diese DOI kann via <a href="http://www.doi.org">www.doi.org</a> 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.



## 6 Registerzeichen

---

Registerzeichen	Erläuterung
AR	Vorverfahren oder sonstige Verfahrensarten
BvA	Verwirkung von Grundrechten
BvB	Verfassungswidrigkeit von Parteien, Art. 21 II GG
BvC	Wahlprüfungsverfahren, Art. 41 II GG
BvD	Anklage des Bundespräsidenten, Art. 61 GG
BvE	Organstreitverfahren, Art. 93 I Nr. 1 GG
BvF	Abstrakte Normenkontrolle, Art. 93 I Nr. 2 GG
BvG	Bund-Länder-Streitigkeiten, Art. 93 Abs. 1 Nr. 3 oder Art. 84 Abs. 4 Satz 2 GG
BvK	Landesverfassungsstreitigkeiten, Art. 99 GG
BvL	Konkrete Normenkontrolle, Art. 100 I GG
BvM	Feststellung der Anwendbarkeit einer Regel des Völkergewohnheitsrechts, Art. 100 II GG iVm Art 25 GG
BvN	Divergenzvorlagen eines Landesverfassungsgerichts zur Auslegung des Grundgesetzes, Art. 100 III GG
BvO	Fortgeltung vorkonstitutionellen Rechts als Bundesrecht, Art. 126 GG
BvP	Sonstige durch Bundesrecht zugewiesene Verfahren, Art. 93 III GG
BvQ	Einstweilige Anordnungen, § 32 BVerfGG
BvR	Verfassungsbeschwerden, Art. 93 Abs. 1 Nr. 4a, und Kommunalverfassungsbeschwerden, Art. 93 Abs. 1 Nr. 4b GG
BvT	Sonstige Verfahren
PBvS	Beendigung des Richteramtes am Bundesverfassungsgericht, § 105 BVerfGG
PBvU	Plenarentscheidungen, § 16 BVerfGG
PBvV	Rechtsgutachten, § 97 BVerfGG a.F.
PKH	Prozesskostenhilfe
Vz	Verzögerungsrügen

---

## 7 PräsidentInnen

### 7.1 Hinweise

Die Daten beziehen sich auf das Amt als PräsidentIn, nicht auf die Amtszeit als VerfassungsrichterIn. Für Roman Herzog ist als Amtsende der Tag vor seinem Antritt als Bundespräsident angegeben. Die erste Tabelle ist maschinenlesbar als CSV zusammen mit dem Source Code hinterlegt. Die zweite Tabelle ist aus der ersten berechnet.

### 7.2 Lebensdaten

Nachname	Vorname	Amtsantritt	Amtsende	Geboren	Gestorben
Höpker-Aschoff	Hermann	1951-09-07	1954-01-15	1883-01-31	1954-01-15
Wintrich	Josef	1954-03-23	1958-10-19	1891-02-15	1958-10-19
Müller	Gebhard	1959-01-08	1971-12-08	1900-04-17	1990-08-07
Benda	Ernst	1971-12-08	1983-12-20	1925-01-15	2009-03-02
Zeidler	Wolfgang	1983-12-20	1987-11-16	1924-09-02	1987-12-31
Herzog	Roman	1987-11-16	1994-09-13	1934-04-05	2017-01-10
Limbach	Jutta	1994-09-14	2002-04-10	1934-03-27	2016-09-10
Papier	Hans-Jürgen	2002-04-10	2010-03-16	1943-07-06	NA
Vofßkuhle	Andreas	2010-03-16	2020-06-22	1963-12-21	NA
Harbarth	Stephan	2020-06-22	NA	1971-12-19	NA

### 7.3 Dienstalter und Lebensalter

Nachname	Vorname	Alter (Amtsantritt)	Alter (Amtsende)	Alter (Tod)
Höpker-Aschoff	Hermann	68	70	70
Wintrich	Josef	63	67	67
Müller	Gebhard	58	71	90
Benda	Ernst	46	58	84
Zeidler	Wolfgang	59	63	63
Herzog	Roman	53	60	82
Limbach	Jutta	60	68	82
Papier	Hans-Jürgen	58	66	NA
Vofßkuhle	Andreas	46	56	NA
Harbarth	Stephan	48	NA	NA

## 8 Vize-PräsidentInnen

### 8.1 Hinweise

Die Daten beziehen sich auf das Amt als Vize-PräsidentIn, nicht auf die Amtszeit als VerfassungsrichterIn. Die erste Tabelle ist maschinenlesbar als CSV zusammen mit dem Source Code hinterlegt. Die zweite Tabelle ist aus der ersten berechnet.

### 8.2 Lebensdaten

Nachname	Vorname	Amtsantritt	Amtsende	Geboren	Gestorben
Katz	Rudolf	1951-09-07	1961-07-23	1895-11-23	1961-07-23
Wagner	Friedrich Wilhelm	1961-12-19	1967-10-18	1894-02-28	1971-03-27
Seuffert	Walter	1967-10-18	1975-11-07	1907-02-04	1989-12-28
Zeidler	Wolfgang	1975-11-07	1983-12-20	1924-09-02	1987-12-31
Herzog	Roman	1983-12-20	1987-11-16	1934-04-05	2017-01-10
Mahrenholz	Ernst Gottfried	1987-11-16	1994-03-24	1929-06-18	NA
Limbach	Jutta	1994-03-24	1994-09-14	1934-03-27	2016-09-10
Henschel	Johann Friedrich	1994-09-29	1995-10-13	1931-06-10	2007-03-19
Seidl	Otto	1995-10-13	1998-02-27	1931-12-11	NA
Papier	Hans-Jürgen	1998-02-27	2002-04-10	1943-07-06	NA
Hassemer	Winfried	2002-04-10	2008-05-07	1940-02-17	2014-01-09
Voßkuhle	Andreas	2008-05-07	2010-03-16	1963-12-21	NA
Kirchhof	Ferdinand	2010-03-16	2018-11-30	1950-06-21	NA
Harbarth	Stephan	2018-11-30	2020-06-22	1971-12-19	NA
König	Doris	2020-06-22	NA	1957-06-25	NA

### 8.3 Dienstalter und Lebensalter

Nachname	Vorname	Alter (Amtsantritt)	Alter (Amtsende)	Alter (Tod)
Katz	Rudolf	55	65	65
Wagner	Friedrich Wilhelm	67	73	77
Seuffert	Walter	60	68	82
Zeidler	Wolfgang	51	59	63
Herzog	Roman	49	53	82
Mahrenholz	Ernst Gottfried	58	64	NA
Limbach	Jutta	59	60	82
Henschel	Johann Friedrich	63	64	75
Seidl	Otto	63	66	NA
Papier	Hans-Jürgen	54	58	NA
Hassemer	Winfried	62	68	73
Voßkuhle	Andreas	44	46	NA
Kirchhof	Ferdinand	59	68	NA
Harbarth	Stephan	46	48	NA
König	Doris	62	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 drei verschiedene klassische linguistische Kennzahlen:

Kennzahl	Definition
Tokens	Eine beliebige Zeichenfolge, getrennt durch whitespace-Zeichen, d.h. ein Token entspricht in der Regel einem »Wort«, kann aber gelegentlich auch 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.

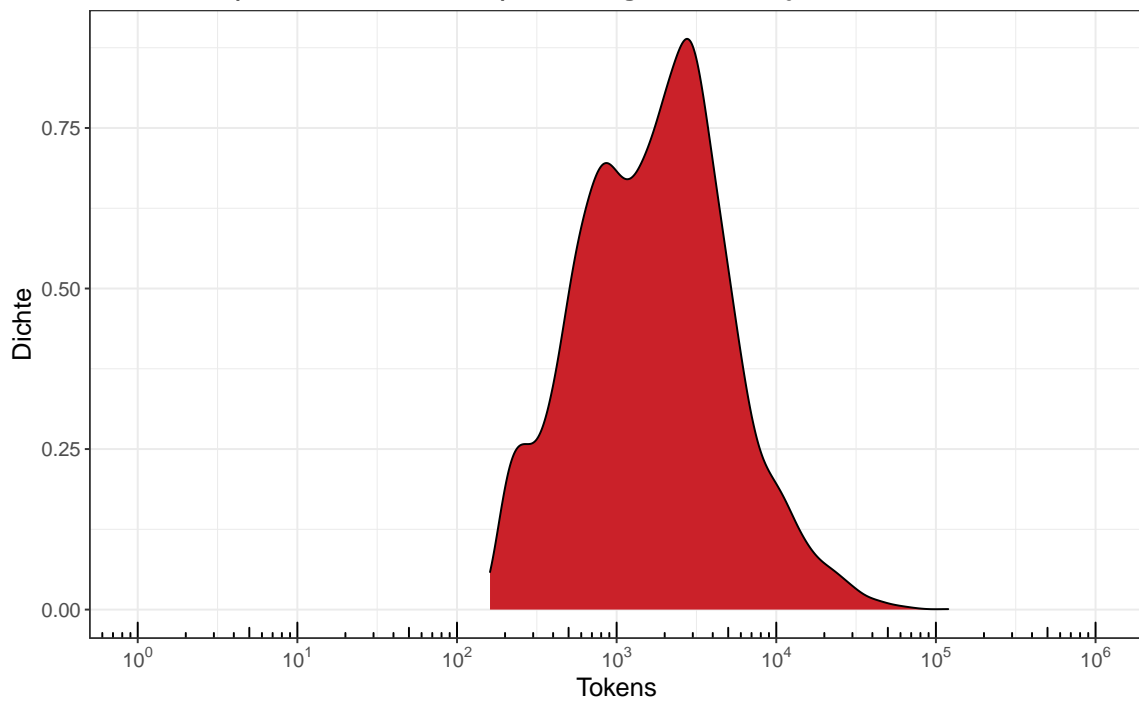
Vorsicht bei der Interpretation: Die x-Achse ist logarithmisch skaliert, d.h. in 10er-Potenzen und damit nicht-linear. Die kleinen Achsen-Markierungen zwischen den Schritten der Exponenten sind eine visuelle Hilfestellung um diese nicht-Linearität zu verstehen.

### 9.2 Werte der Kennzahlen

Kennzahl	Summe	Min	Quart1	Median	Mittel	Quart3	Max
tokens	24,323,364	161	782	1,799	3,102.87	3,491.5	119,512
saetze	1,341,673	6	46	106	171.15	198.5	5,282
typen	NA	91	336	657	876.60	1,104.0	15,930

### 9.3 Verteilung Tokens

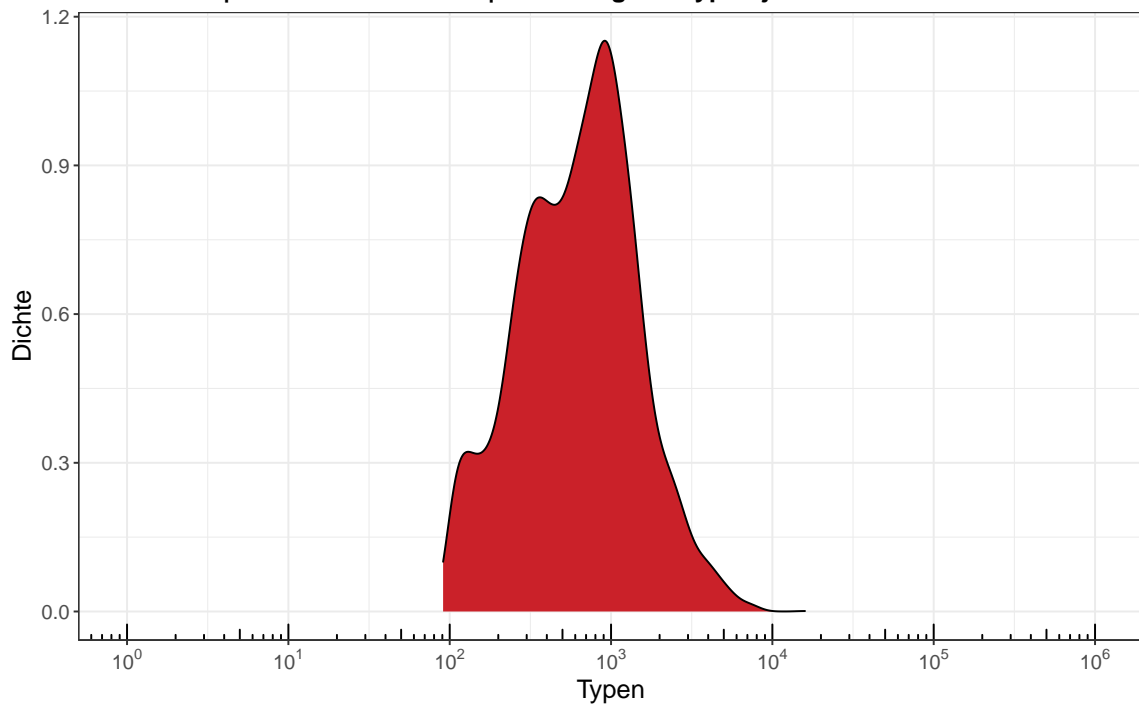
CE-BVerfG | Version 2021-01-08 | Verteilung der Tokens je Dokument



DOI: 10.5281/zenodo.4308215

### 9.4 Verteilung Typen

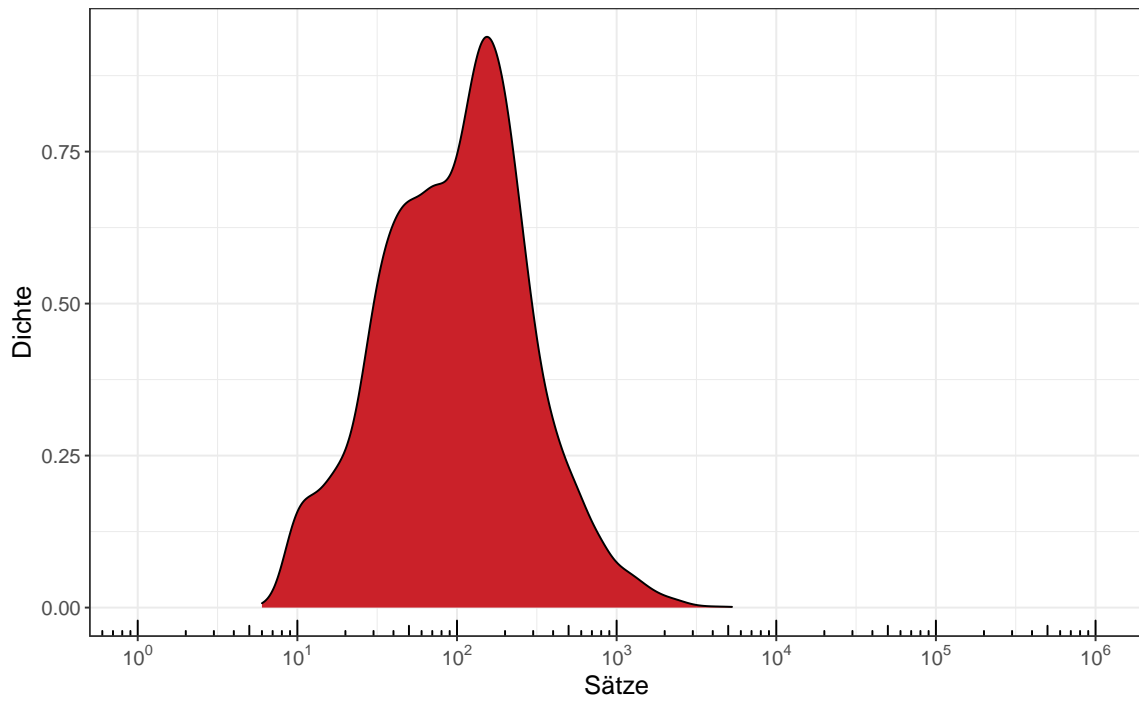
CE-BVerfG | Version 2021-01-08 | Verteilung der Typen je Dokument



DOI: 10.5281/zenodo.4308215

## 9.5 Verteilung Sätze

CE-BVerfG | Version 2021-01-08 | Verteilung der Sätze je Dokument



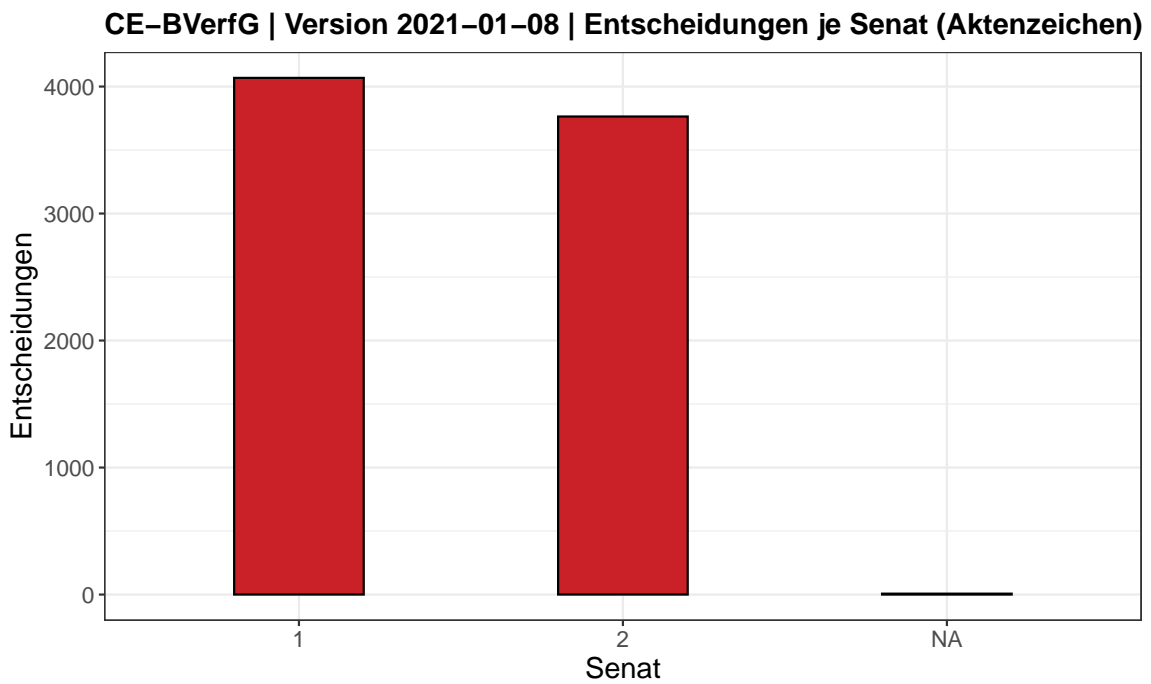
DOI: 10.5281/zenodo.4308215

## 10 Inhalt des Korpus

### 10.1 Zusammenfassung

Variable	Anzahl	Min	Quart1	Median	Mittel	Quart3	Max
entscheidungsjahr	37	1951	2002.0	2008	2008.29	2015	2020
eingangsjahr_iso	44	1951	2001.0	2007	2006.70	2013	2020
band	64	1	106.0	119	120.12	134	152
eingangsnummer	2689	1	301.5	1069	1154.72	1873	3634

### 10.2 Nach Spruchkörper (Aktenzeichen)

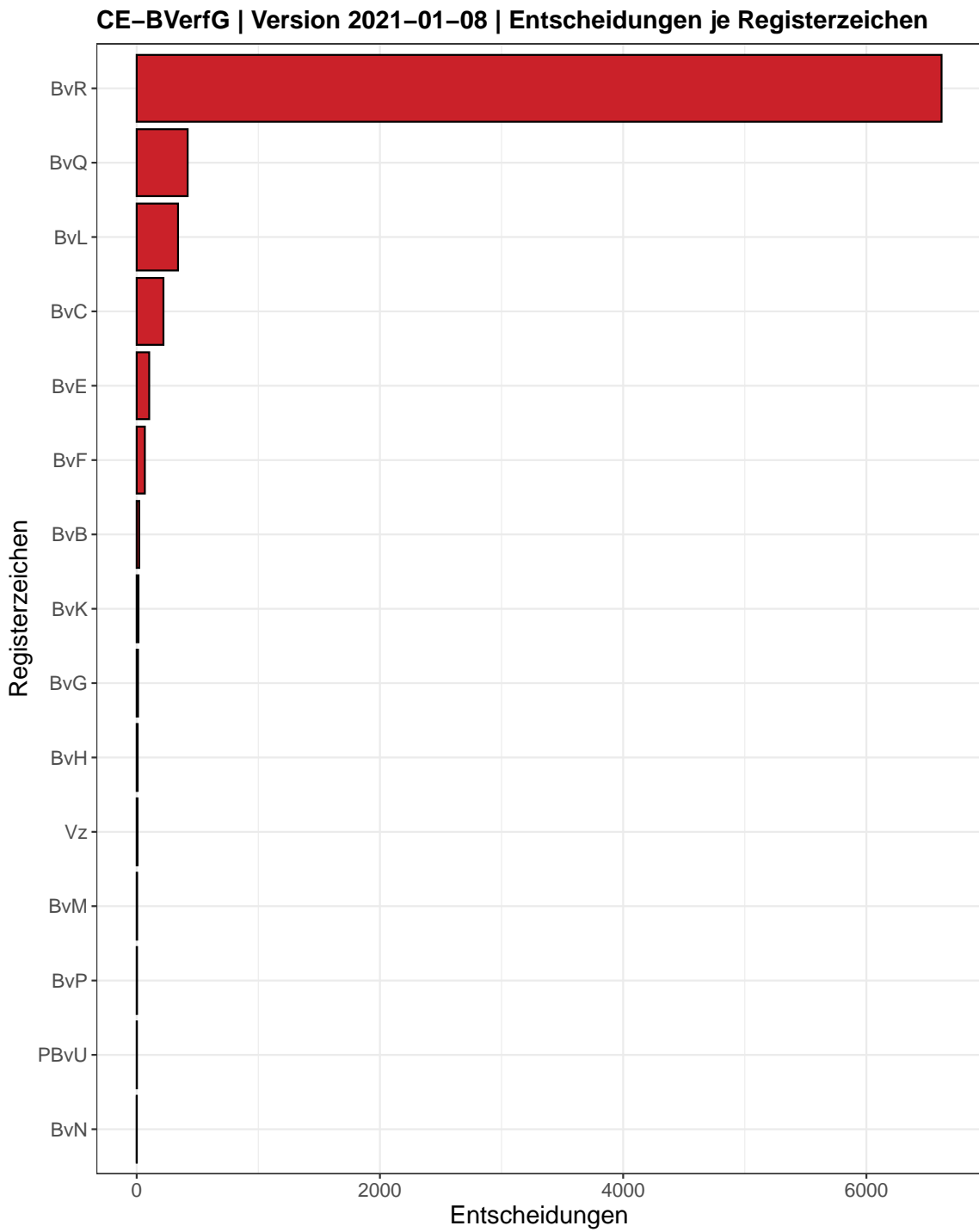


DOI: 10.5281/zenodo.4308215

Senat	Entscheidungen	% Gesamt	% Kumulativ
NA	7	0.09	0.09
1	4068	51.89	51.98
2	3764	48.02	100.00
Total	7839	100.00	100.00

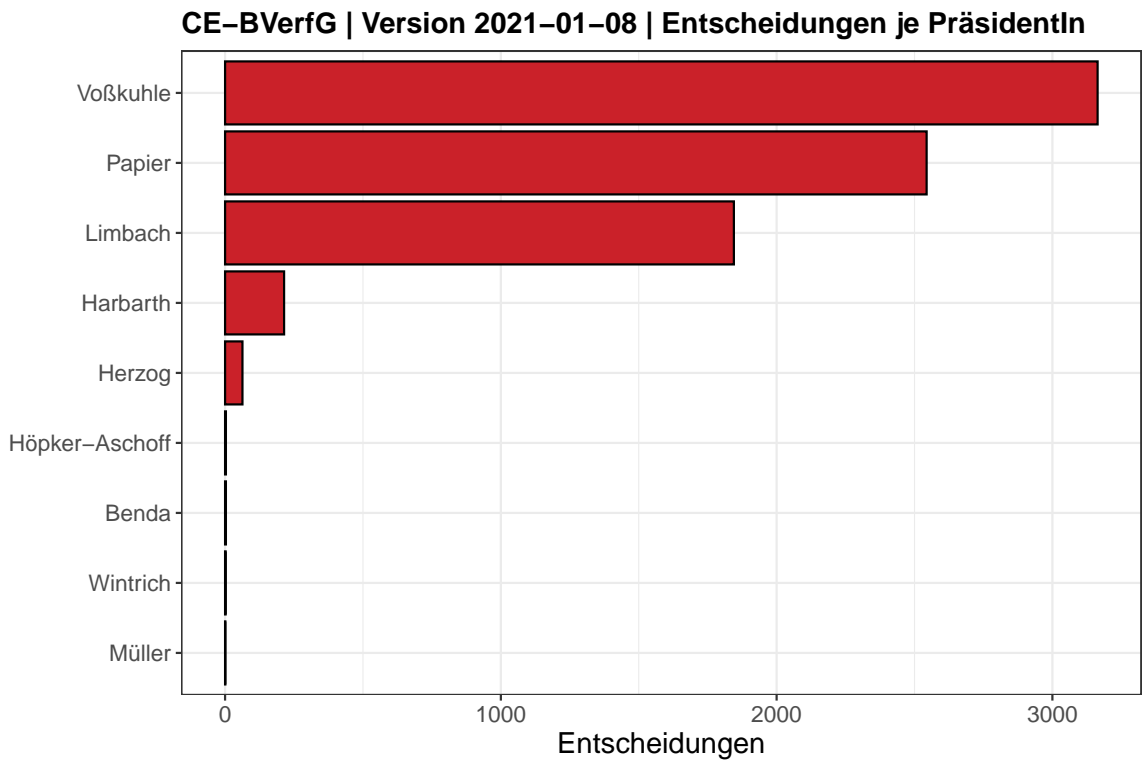


### 10.3 Nach Registerzeichen



Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
BvB	21	0.27	0.27
BvC	220	2.81	3.07
BvE	103	1.31	4.39
BvF	67	0.85	5.24
BvG	11	0.14	5.38
BvH	8	0.10	5.49
BvK	15	0.19	5.68
BvL	340	4.34	10.01
BvM	4	0.05	10.07
BvN	1	0.01	10.08
BvP	3	0.04	10.12
BvQ	419	5.35	15.46
BvR	6618	84.42	99.89
PBvU	2	0.03	99.91
Vz	7	0.09	100.00
Total	7839	100.00	100.00

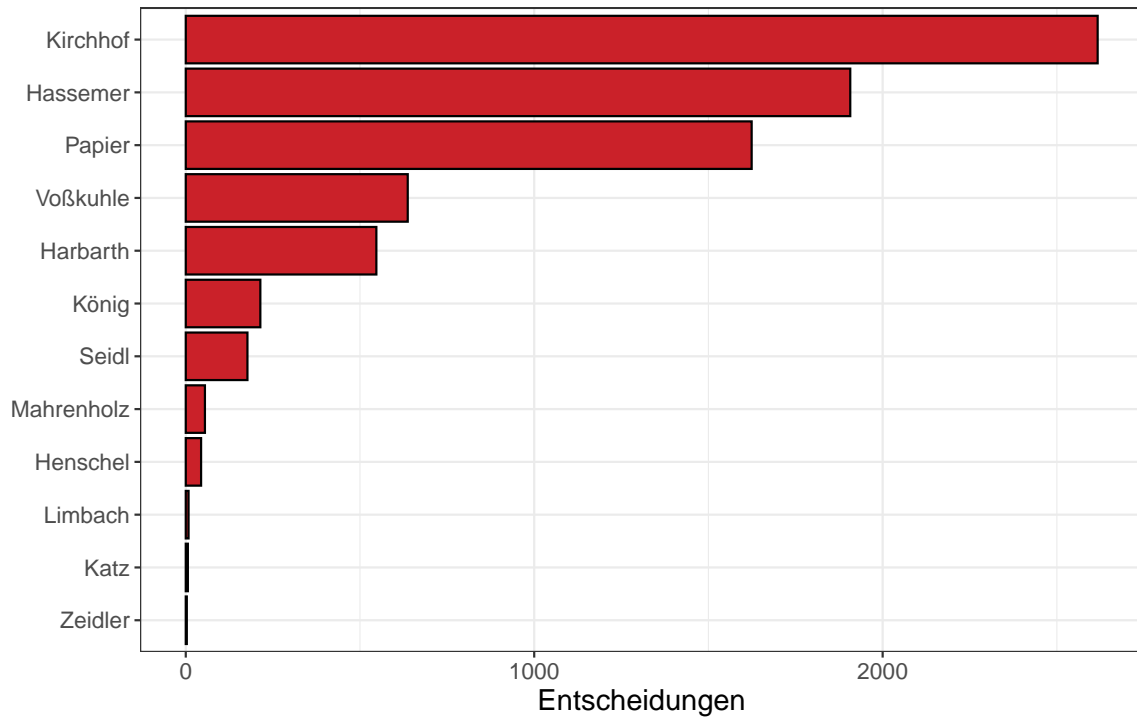
## 10.4 Nach PräsidentIn



PräsidentIn	Entscheidungen	% Gesamt	% Kumulativ
Benda	3	0.04	0.04
Harbarth	214	2.73	2.77
Herzog	63	0.80	3.57
Höpker-Aschoff	3	0.04	3.61
Limbach	1845	23.54	27.15
Müller	1	0.01	27.16
Papier	2544	32.45	59.61
Voßkuhle	3164	40.36	99.97
Wintrich	2	0.03	100.00
Total	7839	100.00	100.00

## 10.5 Nach Vize-PräsidentIn

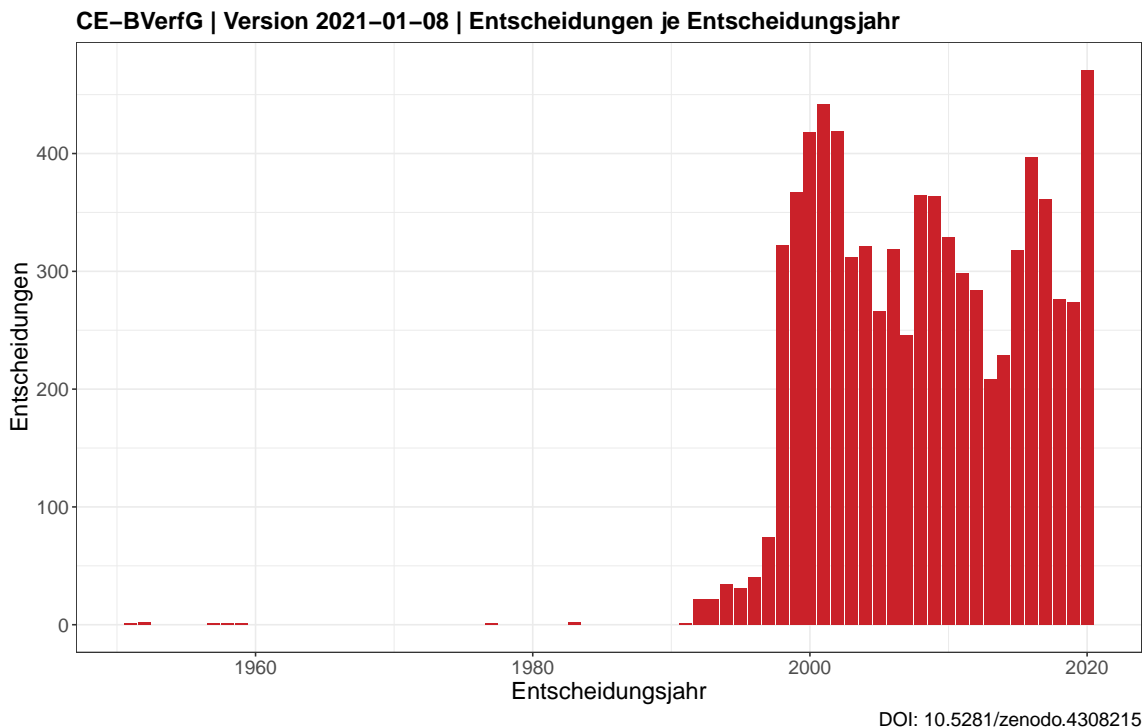
**CE-BVerfG | Version 2021-01-08 | Entscheidungen je Vize-PräsidentIn**



DOI: 10.5281/zenodo.4308215

Vize-PräsidentIn	Entscheidungen	% Gesamt	% Kumulativ
Harbarth	547	6.98	6.98
Hassemer	1907	24.33	31.31
Henschel	44	0.56	31.87
Katz	6	0.08	31.94
Kirchhof	2617	33.38	65.33
König	214	2.73	68.06
Limbach	8	0.10	68.16
Mahrenholz	55	0.70	68.86
Papier	1624	20.72	89.58
Seidl	177	2.26	91.84
Voßkuhle	637	8.13	99.96
Zeidler	3	0.04	100.00
Total	7839	100.00	100.00

## 10.6 Nach Entscheidungsjahr



Jahr	Entscheidungen	% Gesamt	% Kumulativ
1951	1	0.01	0.01
1952	2	0.03	0.04
1957	1	0.01	0.05
1958	1	0.01	0.06
1959	1	0.01	0.08
1977	1	0.01	0.09
1983	2	0.03	0.11
1991	1	0.01	0.13
1992	22	0.28	0.41
1993	22	0.28	0.69
1994	34	0.43	1.12
1995	31	0.40	1.52
1996	40	0.51	2.03
1997	74	0.94	2.97
1998	322	4.11	7.08

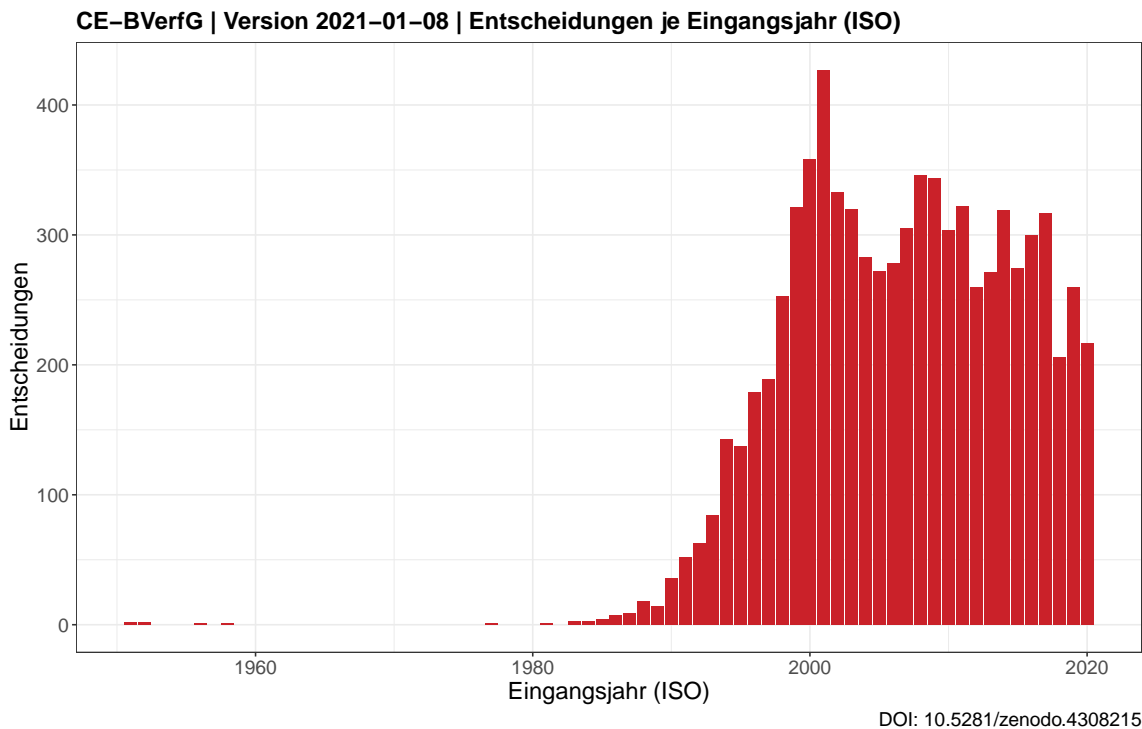
*(continued)*

---

Jahr	Entscheidungen	% Gesamt	% Kumulativ
1999	367	4.68	11.76
2000	418	5.33	17.09
2001	442	5.64	22.73
2002	419	5.35	28.08
2003	312	3.98	32.06
2004	321	4.09	36.15
2005	266	3.39	39.55
2006	319	4.07	43.62
2007	246	3.14	46.75
2008	365	4.66	51.41
2009	364	4.64	56.05
2010	329	4.20	60.25
2011	298	3.80	64.05
2012	284	3.62	67.67
2013	208	2.65	70.33
2014	229	2.92	73.25
2015	318	4.06	77.31
2016	397	5.06	82.37
2017	361	4.61	86.98
2018	276	3.52	90.50
2019	274	3.50	93.99
2020	471	6.01	100.00
Total	7839	100.00	100.00

---

## 10.7 Nach Eingangsjahr (ISO)



Jahr	Entscheidungen	% Gesamt	% Kumulativ
1951	2	0.03	0.03
1952	2	0.03	0.05
1956	1	0.01	0.06
1958	1	0.01	0.08
1977	1	0.01	0.09
1981	1	0.01	0.10
1983	3	0.04	0.14
1984	3	0.04	0.18
1985	4	0.05	0.23
1986	7	0.09	0.32
1987	9	0.11	0.43
1988	18	0.23	0.66
1989	14	0.18	0.84
1990	36	0.46	1.30
1991	52	0.66	1.96

*(continued)*

---

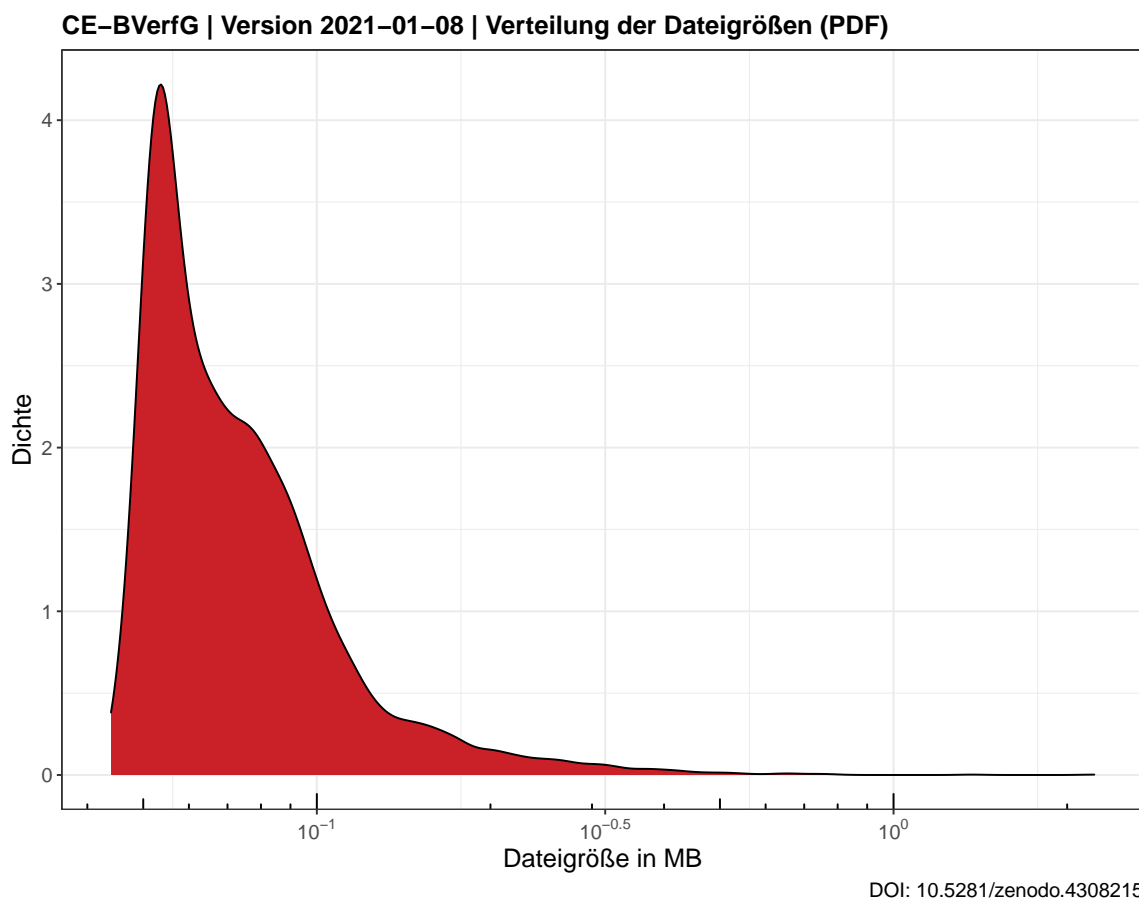
Jahr	Entscheidungen	% Gesamt	% Kumulativ
1992	63	0.80	2.77
1993	84	1.07	3.84
1994	143	1.82	5.66
1995	137	1.75	7.41
1996	179	2.28	9.70
1997	189	2.41	12.11
1998	253	3.23	15.33
1999	321	4.09	19.43
2000	358	4.57	24.00
2001	427	5.45	29.44
2002	333	4.25	33.69
2003	320	4.08	37.77
2004	283	3.61	41.38
2005	272	3.47	44.85
2006	278	3.55	48.40
2007	305	3.89	52.29
2008	346	4.41	56.70
2009	344	4.39	61.09
2010	304	3.88	64.97
2011	322	4.11	69.08
2012	260	3.32	72.39
2013	271	3.46	75.85
2014	319	4.07	79.92
2015	274	3.50	83.42
2016	300	3.83	87.24
2017	317	4.04	91.29
2018	206	2.63	93.92
2019	260	3.32	97.23
2020	217	2.77	100.00

---



## 11 Dateigrößen

Datei	Größe in MB
CE-BVerfG_2021-01-08_DE_ANALYSE.zip	0.53
CE-BVerfG_2021-01-08_DE_CSV_Datensatz.zip	45.72
CE-BVerfG_2021-01-08_DE_CSV_Metadaten.zip	0.28
CE-BVerfG_2021-01-08_DE_PDF_Datensatz.zip	601.98
CE-BVerfG_2021-01-08_DE_TXT_Datensatz.zip	54.22
CE-BVerfG_2021-01-08_Source_Files.zip	0.06



## 12 Signaturprüfung

### 12.1 Allgemeines

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 Hash-Werte in zwei verschiedenen Verfahren berechnet und in einer CSV-Datei dokumentiert. In Phase II wird diese CSV-Datei 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 ist.

Dieses Codebook ist vollautomatisch erstellt und prüft die kryptographisch sicheren SHA3-512 Signaturen (»hashes«) aller ZIP-Archive, sowie die GPG-Signatur der CSV-Datei, welche die SHA3-512 Signaturen enthält. SHA3-512 Signaturen werden durch einen system call zur OpenSSL library auf Linux-Systemen berechnet. Eine erfolgreiche Prüfung meldet »Signatur verifiziert!«. Eine gescheiterte Prüfung meldet »FEHLER!«

### 12.2 Persönliche GPG-Signatur

Die während der Kompilierung des Datensatzes erstellte CSV-Datei mit den Hash-Prüfsummen ist 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

### 12.3 Import: Public Key

```
system2("gpg2", "--import GPG-Public-Key_Fobbe-Data.asc",
        stdout = TRUE,
        stderr = TRUE)
```

```
## [1] "gpg: key 50C41384F44A4E42: \"Sean Fobbe <fobbe-data@posteo.de>\" not
      changed"
## [2] "gpg: Total number processed: 1"
## [3] "gpg:             unchanged: 1"
```

## 12.4 Prüfung: GPG-Signatur der Hash-Datei

```
# CSV-Datei mit Hashes  
print(hashfile)
```

```
## [1] "CE-BVerfG_2021-01-08_KryptographischeHashes.csv"
```

```
# GPG-Signatur  
print(signaturefile)
```

```
## [1] "CE-BVerfG_2021-01-08_FobbeSignatur.gpg"
```

```
# GPG-Signatur prüfen  
testresult <- system2("gpg2",  
                      paste("--verify", signaturefile, hashfile),  
                      stdout = TRUE,  
                      stderr = TRUE)  
  
# Anführungsstriche entfernen um Anzeigefehler zu vermeiden  
testresult <- gsub("'", "", testresult)
```

```
kable(testresult, format = "latex", booktabs=TRUE,  
      longtable=TRUE, col.names = c("Ergebnis"))
```

---

Ergebnis

---

gpg: Signature made Fri 08 Jan 2021 02:02:51 PM CET

gpg: using RSA key FE6FB888F0E5656C1D253B9A50C41384F44A4E42

gpg: Good signature from Sean Fobbe <fobbe-data@posteo.de> [full]

---

## 12.5 Prüfung: SHA3-512 Hashes der ZIP-Archive

```
# Prüf-Funktion definieren
sha3test <- function(filename, sig){
  sig.new <- system2("openssl",
                    paste("sha3-512", filename),
                    stdout = TRUE)
  sig.new <- gsub("^.*\\|= ", "", sig.new)
  if (sig == sig.new){
    return("Signatur verifiziert!")
  }else{
    return("FEHLER!")
  }
}

# Ursprüngliche Signaturen importieren
table.hashes <- fread(hashfile)
filename <- table.hashes$filename
sha3.512 <- table.hashes$sha3.512

# Signaturprüfung durchführen
sha3.512.result <- mcmapply(sha3test, filename, sha3.512, USE.NAMES=FALSE)

# Ergebnis anzeigen
testresult <- data.table(filename, sha3.512.result)
```

```
kable(testresult, format = "latex", booktabs=TRUE,
      longtable=TRUE, col.names = c("Datei", "Ergebnis"))
```

Datei	Ergebnis
CE-BVerfG_2021-01-08_DE_ANALYSE.zip	Signatur verifiziert!
CE-BVerfG_2021-01-08_DE_CSV_Datensatz.zip	Signatur verifiziert!
CE-BVerfG_2021-01-08_DE_CSV_Metadaten.zip	Signatur verifiziert!
CE-BVerfG_2021-01-08_DE_PDF_Datensatz.zip	Signatur verifiziert!
CE-BVerfG_2021-01-08_DE_TXT_Datensatz.zip	Signatur verifiziert!
CE-BVerfG_2021-01-08_Source_Files.zip	Signatur verifiziert!

## 13 Changelog

---

Version	Details
2021-01-08	<ul style="list-style-type: none"><li>• Vollständige Aktualisierung der Daten</li><li>• Veröffentlichung des vollständigen Source Codes</li><li>• Deutliche Erweiterung des inhaltlichen Umfangs des Codebooks</li><li>• Einführung der vollautomatischen Erstellung von Datensatz und Codebook</li><li>• Einführung von Compilation Reports um den Erstellungsprozess exakt zu dokumentieren</li><li>• Einführung von Variablen für Versionsnummer, Concept DOI, Version DOI, ECLI, Entscheidungsnamen, BVerfGE-Band, BVerfGE-Seite, Typ des Spruchkörpers, PräsidentIn, Vize-PräsidentIn und linguistische Kennzahlen (Tokens, Typen, Sätze)</li><li>• Automatisierung und Erweiterung der Qualitätskontrolle</li><li>• Einführung von Diagrammen zur Visualisierung von Prüfergebnissen</li><li>• Einführung kryptographischer Signaturen</li><li>• Alle Variablen sind nun in Kleinschreibung und Snake Case gehalten</li><li>• Variable »Suffix« in »kollision« umbenannt.</li><li>• Variable »Ordinalzahl« in »eingangsnummer« umbenannt.</li></ul>
2020-08-03	<ul style="list-style-type: none"><li>• Vollständige Aktualisierung der Daten</li><li>• Angleichung der Variablen-Namen an andere Datensätze der CE-Serie von <a href="https://zenodo.org/communities/sean-fobbe-data/">https://zenodo.org/communities/sean-fobbe-data/</a></li><li>• Einführung der Variable »Suffix« um weitere Entscheidungen korrekt erfassen zu können; aufgrund der fehlenden Berücksichtigung des Suffix sind die Metadaten von 36 Entscheidungen der Version 2020-06-20 fehlerhaft. Bitte verwenden Sie daher nur die neue Version. Alternativ können Sie die fehlerhaften Dateien (erkennbar an einem dreistelligen Eingangsjahr) aus der Analyse ausschließen oder per Hand korrigieren.</li></ul>
2020-06-20	<ul style="list-style-type: none"><li>• Erstveröffentlichung</li></ul>

---

## 14 Parameter für strenge Replikationen

```
## [1] "OpenSSL 1.1.1i FIPS 8 Dec 2020"
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-redhat-linux-gnu (64-bit)
## Running under: Fedora 32 (Workstation Edition)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib64/libopenblas-r0.3.12.so
##
## locale:
## [1] LC_CTYPE=en_US.utf8      LC_NUMERIC=C
## [3] LC_TIME=en_US.utf8       LC_COLLATE=en_US.utf8
## [5] LC_MONETARY=en_US.utf8   LC_MESSAGES=en_US.utf8
## [7] LC_PAPER=en_US.utf8     LC_NAME=C
## [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.utf8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] parallel stats      graphics grDevices utils      datasets methods
## [8] base
##
## other attached packages:
## [1] magick_2.5.2      quanteda_2.1.2      readtext_0.80      data.table_1.13.4
## [5] scales_1.1.1      ggplot2_3.3.2      doParallel_1.0.16  iterators_1.0.13
## [9] foreach_1.5.1     pdftools_2.3.1     kableExtra_1.3.1   knitr_1.30
## [13] rvest_0.3.6       xml2_1.3.2          httr_1.4.2
##
## loaded via a namespace (and not attached):
## [1] tinytex_0.28      qpdf_1.1            tidyselect_1.1.0   xfun_0.19
## [5] purrr_0.3.4       lattice_0.20-41     colorspace_2.0-0   vctrs_0.3.6
## [9] generics_0.1.0    htmltools_0.5.0    usethis_2.0.0     viridisLite
## [13] yaml_2.2.1        rlang_0.4.9         pillar_1.4.7       glue_1.4.2
## [17] withr_2.3.0       selectr_0.4-2       lifecycle_0.2.0    stringr_1.4.0
## [21] munsell_0.5.0     gtable_0.3.0        codetools_0.2-18  evaluate_0.14
## [25] labeling_0.4.2    curl_4.3            highr_0.8          Rcpp_1.0.5
## [29] RcppParallel_5.0.2 webshot_0.5.2       farver_2.0.3       fs_1.5.0
## [33] fastmatch_1.1-0   stopwords_2.1        askpass_1.1        digest_0.6.27
## [37] stringi_1.5.3     dplyr_1.0.2         grid_4.0.3         tools_4.0.3
## [41] magrittr_2.0.1    tibble_3.0.4        crayon_1.3.4       pkgconfig_2.0.3
## [45] ellipsis_0.3.1    Matrix_1.2-18      rmarkdown_2.5      rstudioapi_0.13
## [49] R6_2.5.0          compiler_4.0.3
```

## Literaturverzeichnis

- Analytics, Revolution, and Steve Weston. 2020. *Iterators: Provides Iterator Construct*. <https://github.com/RevolutionAnalytics/iterators>.
- Benoit, Kenneth, and Adam Obeng. 2020. *Readtext: Import and Handling for Plain and Formatted Text Files*. <https://github.com/quanteda/readtext>.
- 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, Jiong Wei Lua, Jouni Kuha, and William Lowe. 2020. *Quanteda: Quantitative Analysis of Textual Data*. <https://quanteda.io>.
- Corporation, Microsoft, and Steve Weston. 2020. *DoParallel: Foreach Parallel Adaptor for the Parallel Package*. <https://CRAN.R-project.org/package=doParallel>.
- Dowle, Matt, and Arun Srinivasan. 2020. *Data.table: Extension of ‘Data.frame’*. <https://CRAN.R-project.org/package=data.table>.
- Ooms, Jeroen. 2020. *Pdftools: Text Extraction, Rendering and Converting of Pdf Documents*. <https://CRAN.R-project.org/package=pdfutils>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Revolution Analytics, and Steve Weston. n.d. *Foreach: Provides Foreach Looping Construct*.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2020a. *Httr: Tools for Working with Urls and Http*. <https://CRAN.R-project.org/package=httr>.
- . 2020b. *Rvest: Easily Harvest (Scrape) Web Pages*. <https://CRAN.R-project.org/package=rvest>.
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2020. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2>.
- Wickham, Hadley, Jim Hester, and Jeroen Ooms. 2020. *Xml2: Parse Xml*. <https://CRAN.R-project.org/package=xml2>.
- Wickham, Hadley, and Dana Seidel. 2020. *Scales: Scale Functions for Visualization*. <https://CRAN.R-project.org/package=scales>.
- 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. <http://www.crcpress.com/product/isbn/9781466561595>.

———. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.

———. 2020. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://yihui.org/knitr/>.

Zhu, Hao. 2020. *KableExtra: Construct Complex Table with Kable and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.