

German Index of Socioeconomic Deprivation (GISD)

Robert Koch-Institut | RKI

Nordufer 20
13353 Berlin

Niels Michalski, Marvin Reis, Fabian Tetzlaff, Enno Nowossadeck und Jens Hoebel

(Projektleitung)

FG 28 | Soziale Determinanten der Gesundheit

Zitieren

Michalski, Niels; Soliman, Omar; Marvin Reis; Fabian Tetzlaff; Enno Nowossadeck; Jens Hoebel (2024): German Index of Socioeconomic Deprivation (GISD), Berlin: Zenodo. DOI: [10.5281/zenodo.10457601](https://doi.org/10.5281/zenodo.10457601)

Informationen zum Forschungskontext

Im Datensatz "German Index of Socioeconomic Deprivation (GISD)" werden die Daten des gleichnamigen Indexes bereitgestellt und dokumentiert. Er dient als Maß relativer regionaler sozioökonomischer Benachteiligung. Der German Index of Socioeconomic Deprivation (GISD) wird seit 2017 am Robert Koch-Institut erstellt, um regionale sozioökonomische Ungleichheiten in der Gesundheit sichtbar zu machen und Ansatzpunkte zur Erklärung regionaler Unterschiede in der Gesundheit aufzeigen zu können. Die Indikatoren des GISD werden so gewählt, dass sie eine Analogie zum in der Sozialepidemiologie etablierten individuellen [sozioökonomischen Status \(SES\)](#) zulassen, der die Dimensionen Bildung, Beruf und Einkommen zusammenfasst. Durch diese Konzeption wird es möglich mit dem GISD sozioökonomische Unterschiede in den Gesundheitschancen, Krankheits- und Sterberisiken in Deutschland auch anhand von Daten untersuchen zu können, die auf individueller Ebene keine Information zum sozioökonomischen Status enthalten. Der Begriff regionale Deprivation wird hier in Anlehnung an den Begriff der sozialen Deprivation verwendet und bezeichnet den Grad der Benachteiligung, der aus einem relativen Mangel an sozioökonomischen Ressourcen und aus einem hohen Niveau sozialräumlicher Belastungen in der jeweiligen Wohnregion resultiert, der Einschränkungen der gesellschaftlichen Teilhabechancen zur Folge hat.

Für die Referenzierung des GISD und seiner Methodik wird folgende Publikation im Journal of Health Monitoring empfohlen, in der eine ausführliche Einführung in den GISD erfolgt. Einige Textbausteine dieser Onlinedokumentation wurden dieser Publikation entnommen:

Michalski N, Reis M, Tetzlaff F, Herber M, Kroll LE, Hövener C, Nowossadeck E, Hoebel J (2022) German Index of Socioeconomic Deprivation (GISD): Revision, Aktualisierung und Anwendungsbeispiele. J Health Monit 7(S5): 2–24. DOI: [10.25646/10640](https://doi.org/10.25646/10640)

Die englische Version ist verfügbar unter:

Michalski N, Reis M, Tetzlaff F, Herber M, Kroll LE et al. (2022) German Index of Socioeconomic Deprivation (GISD): Revision, update and applications. J Health Monit 7(S5): 2–23. DOI [10.25646/10641](https://doi.org/10.25646/10641)

Die Begleitpublikation zur früheren Version des GISD findet sich hier:

Kroll LE, Schumann M, Hoebel J et al. (2017) Regionale Unterschiede in der Gesundheit – Entwicklung eines sozioökonomischen Deprivationsindex für Deutschland. J Health Monit 2(2):103-120 DOI: [10.25646/2573.2](https://doi.org/10.25646/2573.2)

Administrative und organisatorische Angaben

Der "German Index of Socioeconomic Deprivation (GISD)" dient im Rahmen der epidemiologischen Forschung und Gesundheitsberichterstattung am Robert Koch-Institut zur Bearbeitung der Themenschwerpunkte [regionale Unterschiede in der Gesundheit](#) und [soziale Ungleichheit der Gesundheitschancen und Erkrankungsrisiken](#).

Die Federführung der Entwicklung, Betreuung und Primäranwendung liegt im [Fachgebiet 28 | Sozialen Determinanten der Gesundheit](#). Autoren des aktuellen Datensatzes sind [Niels Michalski](#) und [Marvin Reis](#). Die Projektleitung liegt bei [Jens Hoebel](#). Die [Ursprungsversion des GISD](#) wurde von [Lars E. Kroll](#) erstellt. Aktualisierungen und Revisionen des GISD erscheinen in regelmäßigen Abständen. Fragen zur Generierung und Verwendung des GISD können an GISD@rki.de gerichtet werden.

Die Veröffentlichung der Daten, die Datenkuration sowie das Qualitätsmanagement der (Meta-)Daten erfolgt durch das Fachgebiet [MF 4 | Informations- und Forschungsdatenmanagement](#). Fragen zum Datenmanagement können an das Open Data Team des Fachgebiets MF4 gerichtet werden OpenData@rki.de.

Zielkriterien und Datengrundlage

Beschreibung der Zielkriterien

Die Erstellung des GISD erfolgte mit dem Ziel, ein Maß für die relative sozioökonomische Schlechterstellung von Regionen in Deutschland zu entwickeln. Der Index sollte eine Analogie zum sozioökonomischen Status (SES) auf Individualebene zulassen und auch Entwicklungen über die Zeit abbilden können. Der Index sollte sozioökonomische Unterschiede dabei möglichst feingliedrig, d.h. auf möglichst kleinräumiger regionaler Ebene, abbilden. Primärer Bezugsrahmen des Indexes sollte das gesamte Bundesgebiet sein, um Gesundheitsdaten mit Regionalbezug aus unterschiedlichen Quellen analysieren und für Deutschland repräsentative

Aussagen treffen zu können. Als Datengrundlage sollten Sekundärdaten aus Quellen verwendet werden, die frei zugänglich sind und deren datenschutz- und urheberrechtliche Verwendung unproblematisch ist.

Regionale Ebenen

Gemäß dieser Kriterien wird der GISD auf Ebene der Gemeinden jahresweise für den Zeitraum 1998 bis 2019 generiert und für die unterschiedliche Raumabgrenzungen (Gemeindeverbände, Kreis, Raumordnungsregionen, NUTS-2) bevölkerungsgewichtet aggregiert. Auf jeder räumlichen Ebene wird jeder Region des Gebietsstands ein Wert zugeordnet.

Ebene	Anzahl (Stand 31.12.2019)
Gemeinden	10799
Gemeindeverbände und verbandsfreie Gemeinden (GVB)	4411
Kreise und kreisfreie Städte (Kreise)	401
Raumordnungsregionen (ROR)	96
NUTS-2	38

Datengrundlage

Für die Generierung des GISD werden bevorzugt Daten aus der Datenbank [INKAR](#) (Indikatoren und Karten zur Raum- und Stadtentwicklung) des Bundesinstituts für Bau-, Stadt- und Raumforschung (BBSR) verwendet. Bei INKAR handelt es sich um einen interaktiven Online-Atlas mit regionalstatistischen Informationen zu den Themen Bevölkerungsentwicklung, Arbeitsmarkt, Bildung, Wirtschaft, Wohnen, Verkehr und Umwelt. In der Datenbank sind rund 600 Indikatoren für verschiedene regionale Ebenen hinterlegt, so dass Vergleiche zwischen europäischen Regionen, Bundesländern, Kreisen, Gemeindeverbänden und verbandsfreien Gemeinden ermöglicht werden. Die Untergliederung der regionalen Einheiten basiert auf der [administrativen Gliederung Deutschlands](#). Die Zeitreihen der Statistiken reichen zum Teil bis ins Jahr 1995 zurück. Die INKAR-Daten bieten durch ihre öffentliche Verfügbarkeit und gute Dokumentation ein hohes Maß an Transparenz. Zudem werden die Daten über die Zeit harmonisiert und auf die aktuellen Gebietsstände bezogen, so dass zeitliche Trends analysiert werden können. Aufgrund der genannten Eigenschaften werden Daten aus der INKAR-Datenbank gegenüber anderen Datenquellen bevorzugt. Die Zeitreihen für zwei Indikatoren wurden von der [Statistik der Bundesagentur für Arbeit](#) bezogen.

Auswahl der Indikatoren

Der Index greift im Wesentlichen auf die Indikatoren zurück, deren Auswahl in einer früheren

Version des GISD auf Grundlage der Ergebnisse einer systematischen Literaturrecherche in den Literaturdatenbanken von PubMed und Google Scholar getroffen wurde [Kroll et al 2017](#). Die Indikatoren sollten außerdem drei weitere Kriterien erfüllen: Erstens die regionale Auflösung der Daten sollte möglichst fein sein. Sie sollten mindestens auf Kreisebene, idealerweise auf Ebene der Gemeindeverbände und verbandsfreien Gemeinden (GVB) vorliegen. Zweitens sollten die Indikatoren möglichst über einen Zeitraum von derzeit rückblickend 20 Jahren verfügbar und auf aktuelle Gebietsstände bezogen sein. Drittens sollten die Indikatoren auf einer der Teildimensionen mit ausreichender Höhe laden, um eine Zusammenfassung zu einer Dimension zu rechtfertigen. Für die aktuelle Version des GISD wurden die derzeit in der INKAR-Datenbank bereitgestellten Indikatoren auf ihre Eignung hin überprüft. Hierbei stellte sich heraus, dass alle acht ursprünglich verwendeten Indikatoren auch weiterhin in der INKAR-Datenbank enthalten sind. Als problematisch erwies sich der Indikator „Anteil der sozialversicherungspflichtig Beschäftigten mit akademischem Berufsabschluss an allen sozialversicherungspflichtig Beschäftigten“, da dieser in der Datenbankversion von INKAR 2021 nur noch auf den Arbeitsort und nicht mehr auf den Wohnort bezogen bereitgestellt wurde. Der Bezug zum Wohnort ist für den Indikator alternativlos, damit er als Näherungswert für das Bildungsniveau in der regionalen Wohnbevölkerung gelten kann. Die Zeitreihen für den wohnortbezogenen Indikator wurden daher für die aktuelle GISD-Revision direkt von der Statistik der Bundesagentur für Arbeit bezogen. In diesem Zuge wurde auch der Indikator „Anteil sozialversicherungspflichtig Beschäftigter ohne Berufsabschluss an allen sozialversicherungspflichtig Beschäftigten“ neu in die Liste der Indikatoren aufgenommen. Dadurch kann die Bildungsdimension in der vorliegenden Revision wie die Einkommens- und Beschäftigungsdimension durch drei Indikatoren repräsentiert werden. Um der mangelnden Analogie der Indikatoren der Beschäftigungsdimension mit denen des Berufsstatus im sozioökonomischen Status (SES) gerecht zu werden, weicht die Bezeichnung dieser GISD-Dimension von der des SES ab. In folgender Tabelle sind die Dimensionen mit ihren Indikatoren und der jeweiligen Ursprungsquelle der Daten dargestellt:

Dimension	Merkmal	Indikator	Quelle
Bildung	Beschäftigte mit (Fach-)Hochschulabschluss	Anteil der SV Beschäftigten am Wohnort mit akademischem Abschluss an den SV Beschäftigten am Wohnort in %	Statistik der Bundesagentur für Arbeit
Bildung	Beschäftigte ohne Abschluss	Anteil der SV Beschäftigten am Wohnort ohne Berufsabschluss an den SV Beschäftigten am Wohnort in %	Statistik der Bundesagentur für Arbeit
	Schulabgänger-	Anteil der Schulabgängerinnen	Statistik der

Bildung	innen und - abgänger ohne Abschluss	und -abgänger ohne Hauptschulabschluss an den Schulabgängerinnen und -abgängern in %	allgemein- bildenden Schulen des Bundes und der Länder
Beschäf- tigung	Arbeitslosigkeit	Anteil der Arbeitslosen an den Einwohnerinnen und Einwohnern im erwerbsfähigen Alter	Statistik der Bundesagentur für Arbeit
Beschäf- tigung	Beschäftigtenquote	SV Beschäftigte am Wohnort je 100 Einwohnerinnen und Einwohner im erwerbsfähigen Alter	Statistik der Bundesagentur für Arbeit
Beschäf- tigung	Bruttolohn und - gehalt	Monatlicher Bruttoverdienst der Arbeitnehmerinnen und Arbeitnehmer in Euro	Volkswirtschaftliche Gesamtrechnung der Länder
Einkommen	Haushalts- nettoeinkommen	Durchschnittliches Haushalts- einkommen in € je Einwohnerinnen und Einwohner	Volkswirtschaftliche Gesamtrechnung der Länder
Einkommen	Schuldnerquote	Private Schuldnerinnen und Schuldner je 100 Einwohnerinnen und Einwohner 18 Jahre und älter in %	Schuldneratlas Deutschland des Verbands der Vereine Creditreform e.V.
Einkommen	Steuereinnahmen	Einkommensteuer in € je Einwohnerinnen und Einwohner	Realsteuervergleich des Bundes und der Länder

Quellen: [INKAR 2021](#), [Statistik der Bundesagentur für Arbeit](#)

GVB = Gemeindeverbände und verbandsfreie Gemeinden

SV = sozialversicherungspflichtig

*Datenquelle für die Anteile der sozialversicherungspflichtig Beschäftigten ohne Abschluss und mit akademischem Abschluss ist die [Statistik der Bundesagentur für Arbeit](#). Daten für die Jahre 2013 bis 2019 sind dort frei verfügbar. Daten für die Jahre 2001 bis 2011 wurden direkt über die Statistik der Bundesagentur für Arbeit bezogen.

**Für die Arbeitslosigkeit und die Beschäftigtenquote standen Werte für die GVB-Ebene erst ab 2001 zur Verfügung. Für die Jahre 1998 bis 2001 wurden den GVB die Werte der Kreisebene zugeordnet.

[INKAR \(2021\) Indikatoren und Karten zur Raum- und Stadtentwicklung. Ausgabe 2021.](#)
Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR), Bonn.

[Statistik der Bundesagentur für Arbeit \(2022\) Tabellen, Beschäftigte nach Berufen \(KIdB 2010\) \(Quartalszahlen\), Nürnberg.](#)

Datenaufbereitung

Zeitreihenimputation und Datenharmonisierung

Für drei der neun Indikatoren sind Daten auf der Ebene der Gemeindeverbände und verbandsfreien Gemeinden (GVB) verfügbar, so dass der Index auch Unterschiede in der sozioökonomischen Deprivation zwischen GVB innerhalb von Kreisen feststellen kann. Für die Indikatoren Arbeitslosigkeit und Beschäftigtenquote lagen die Daten für die Jahre 1998 bis 2000 nur auf der Kreisebene vor, so dass die Variation zwischen GVB für diese frühen Jahre nochmals geringer ausfällt. Für die Bildungsindikatoren „Beschäftigte ohne Abschluss“ und „Beschäftigte mit (Fach-)Hochschulabschluss“ gibt es aufgrund einer Umstellung im Meldeverfahren zur Sozialversicherung keine Werte für das Jahr 2012 ([Statistik der Bundesagentur für Arbeit](#)). Um diese Zeitreihenlücke zu füllen wurden die Daten für 2011 auf 2012 übertragen. Für einige Indikatoren beginnen die Zeitreihen nicht im Basisjahr 1998. Für diese Indikatoren wurden fehlende Werte auf Grundlage der vorhandenen Zeitreihen geschätzt (lineare Random-Intercept-Modelle für Zeitreihen). Das betrifft die Jahre vor 2004 und insbesondere die Jahre 1998 bis 2000, für die drei bis fünf Indikatoren auf diese Weise ersetzt wurden. Von 2001 bis 2003 mussten nur fehlende Werte eines Indikators geschätzt werden.

Um die Informationen der Rohdaten als Zeitreihe verwenden zu können und für die Hauptkomponentenanalyse aufzubereiten, wurden einige Indikatoren um Artefakte der Erhebung und statistische Artefakte bereinigt: 1) Die Indikatoren mit währungsbasierten, nach oben offenen Skalen (Steuereinnahmen, Bruttolohn und mittleres Haushaltsnettoeinkommen) wurden kaufkraftbereinigt und logarithmiert. 2) Für die Indikatoren „Beschäftigte ohne Abschluss“ und „Beschäftigte mit (Fach-)Hochschulabschluss“ führte die bereits erwähnte Umstellung im Meldeverfahren zur Sozialversicherung zu einer Veränderung der Niveaus der Zeitreihe. Die durchschnittliche Veränderung des Niveauparameters durch die Umstellung des Meldeverfahrens wurde statistisch identifiziert und die Zeitreihe vor der Umstellung auf das Niveau nach der Umstellung angepasst. 3) Der Indikator „Schulabgängerinnen und -abgänger ohne Abschluss“ wurde um Einflüsse der G8-Reformen bei den jährlichen Zahlen der Schulabgängerinnen und -abgänger bereinigt. Diese Korrektur war notwendig, da die Schulabschlussjahrgänge in den Jahren, in denen die ersten Schülerinnen und Schüler im neu eingeführten G8-System nach zwölf Jahren ihr Abitur machen, doppelt besetzt sind und der

Anteil der Schulabgängerinnen und -abgänger ohne Abschluss deshalb aufgrund der insgesamt höheren Abschlusszahlen geringer ausfällt. Um diese Artefakte herauszurechnen wurde der statistische Effekt der G8-Reformen mit den Daten regressionsanalytisch geschätzt und für die betroffenen Bundesländer in den betreffenden Jahren abgezogen. 4) Bei der Betrachtung des Indikators Anteil der Beschäftigten ohne Abschluss fiel eine bimodale Verteilung auf, die auf den deutlich geringeren Anteil an Beschäftigten ohne Abschluss in den ostdeutschen Bundesländern zurückzuführen ist. Inhaltlich sollte der Indikator als Proxy für den Bildungsstand der Bevölkerung dienen. Historische gewachsene Unterschiede zwischen den Landesteilen sowie die Einbindung der DDR-Bevölkerung ins Erwerbssystem der BRD sorgen in dieser Frage allerdings für unterschiedliche Vergleichsmaßstäbe. Um für diese historisch geprägten und demografischen Unterschiede zu korrigieren, wurde die mittlere Differenz der entsprechenden Anteile zwischen ost- und westdeutschen Bundesländern bestimmt und den Regionen in den ostdeutschen Bundesländern zugeschlagen, wodurch eine unimodale Verteilung der Werte erzeugt wurde.

Indexbildung

Die Bestimmung der Indexwerte erfolgte in drei Schritten. In einem ersten Schritt wurden separate Hauptkomponentenanalysen für jede der drei Teildimension durchgeführt. Aus diesen wurden in einem zweiten Schritt Werte für die Teildimensionen über das Faktor-Scoring-Verfahren generiert. In einem dritten Schritt wurden die Werte für die Teildimensionen normalisiert und zum GISD-Score aufsummiert. Für die Hauptkomponentenanalysen wurden die Daten der GISD-Indikatoren für die Jahre 2001-2019 gepoolt. Dadurch konnte neben der Variation zwischen den regionalen Einheiten auch die Variation über die Zeit genutzt werden, um die Zusammenhänge zwischen den Indikatoren zu bestimmen. In die Hauptkomponentenanalyse wurden nur Daten ab 2001 eingeschlossen, um starken Verzerrungen der Gewichte durch imputierte Werte vorzubeugen. Die Hauptkomponentenanalysen bestätigten für jede Teildimension eine eindimensionale Struktur. Die Faktorladungen, die als Korrelation der Indikatoren mit der Hauptkomponente gelesen werden können, erreichten zufriedenstellende bis gute Werte:

Dimension	Variable	Faktorladung	Anteil am Teilscore	Anteil GSD
Bildung	Beschäftigte mit akademischem Abschluss	-0,732	34,1	11,4
Bildung	Beschäftigte ohne Abschluss (adj.)	0,771	37,8	12,6
Bildung	Schulabgängerinnen und -abgänger ohne Abschluss (adj.)	0,663	28,0	9,3
Beschäftigung	Beschäftigtenquote	-0,640	23,1	7,7
Beschäftigung	Arbeitslosigkeit	0,841	39,9	13,3
Beschäftigung	Bruttoverdienst (log.)	-0,810	37,0	12,3
Einkommen	Einkommensteuer (log.)	-0,911	40,6	13,5
Einkommen	Haushaltseinkommen (log.)	-0,921	41,5	13,8
Einkommen	Schuldnerquote	0,607	18,0	6,0

log. = logarithmiert,
adj. = adjustiert

*Darstellung der Faktorladungen auf der ersten Hauptkomponente für die einzelnen Dimensionen. Eigenwerte der ersten Komponenten: $\eta(\text{Bildung})=1,6$; $\eta(\text{Beschäftigung})=1,8$; $\eta(\text{Einkommen})=2,0$. Die Eigenwerte der zweiten und dritten Komponenten liegen jeweils unter 0,8. Die Korrelationen zwischen den Scores der Teildimensionen für das Jahr 2019: $r(\text{Beschäftigung}|\text{Einkommen}) = 0,66$; $r(\text{Beschäftigung}|\text{Bildung}) = 0,52$; $r(\text{Einkommen}|\text{Bildung}) = 0,69$.

Im zweiten Schritt wurden Werte (Scores) für die Teildimensionen mit Hilfe des Faktor-Scorings für jede Gemeinde im Datensatz berechnet. Auch Werte für die Jahre 1998 bis 2000 wurden dabei berücksichtigt. Die Faktorwerte der drei Teildimensionen wurden anschließend jahresweise jeweils auf einen Wertebereich zwischen 0 (niedrigste Deprivation) und 1 (höchste Deprivation) normiert. Danach wurden die drei Teilscores aufaddiert, so dass jede Teildimension mit einem Gewicht von etwa einem Drittel in den Gesamt-Index einging. Die so entstandenen GSD-Scores der Gemeinden wurden anschließend jeweils für höhere regionale Einheiten (GVB, Landkreise und kreisfreie Städte, Raumordnungsregionen und die statistischen Regionen gemäß der amtlichen europäischen Statistik NUTS-2) bevölkerungsgewichtet aggregiert. Die

Berechnung der GISD-Scores für die höheren regionalen Ebenen mittels bevölkerungsgewichteter Aggregation sorgt dafür, dass zur Berechnung der GISD-Scores für jede räumliche Ebene die gleiche Gewichtung der Indikatoren zur Anwendung kommt. Ausgehend von der Gemeindeebene werden auf diese Weise auch Werte für weitere Raumbezüge wie Postleitzahlgebiete generiert. Die Werte wurden auf den jeweiligen Ebenen jahresweise so normiert, dass die sozioökonomische Deprivation der Regionen Werte zwischen 0 (niedrigste Deprivation) und 1 (höchste Deprivation) annehmen kann. Zudem wurden die Einheiten der genannten Raumebenen für die weiteren Analysen gemäß der Verteilung ihrer Indexwerte für jedes Jahr in fünf Gruppen von je zwanzig Prozent (Quintile, Fünftel) eingeteilt, wobei jeweils das unterste Fünftel „niedrige“ und das oberste Fünftel „hohe“ sozioökonomische Deprivation kennzeichnet.

Datenbereitstellung

Im Repositorium werden die GISD-Daten bestehend aus den GISD-Scores für die jeweiligen regionalen Ebenen und Jahre bereitgestellt. Im Ordner "Bund" befinden sich die GISD-Daten mit Bezugsrahmen Bundesebene. Im Ordner "Bundesland" sind die GISD-Daten bezogen auf die jeweiligen Bundesländer mit Ausnahme der Stadtstaaten Bremen, Hamburg und Berlin abgelegt. Innerhalb der Stadtstaaten kann bisher nicht regional differenziert werden. Das aktuelle Release entspricht dem [GISD Release 2022 v1.1](#) des [Archivs](#).

[GISD_Release_aktuell/Bund](#)

[GISD_Release_aktuell/Bundesland](#)

Die Dateiname sind nach dem Schema `GISD_Bezugsrahmen_Raumabgrenzung.csv` bezeichnet. Die Sequenz "Bezugsrahmen" repräsentiert ob als Bezugsrahmen die Bundesebene "Bund" oder ein bestimmtes Bundesland "Name des Bundeslandes" ausgewählt werden soll. Die Sequenz "Raumabgrenzung" markiert die "regionale Ebene" auf der der GISD verwendet werden soll.

Beispiel: Die Datei `GISD_Sachsen_Gemeinde.csv` enthält die Daten des GISD auf Gemeindeebene mit dem Bezugsrahmen Sachsen.

Variablen und Variablenausprägungen

Die Datensätze enthalten folgende Merkmale:

- Information zur regionalen Einheit (IDs und Namen)
- GISD-Score
- GISD-Quintil für den jeweiligen Bezugsrahmen
- GISD-Perzentil für den jeweiligen Bezugsrahmen
- GISD-Kategorien
- Beobachtungsjahr

Die GISD-Daten enthalten die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen:

Variable	Typ	Ausprägung	Beschreibung
gemeinde_id, gvb_id, kreis_id, nuts2_id, ror_id, plz2, plz3, plz4, plz5	String	01001 DE01 ...	ID der regionalen Einheit basierend auf dem amtlichen Regionalschlüssel für Gemeinden, Gemeindeverbände, Stadt- und Landkreise, NUTS-2, Raumordnungsregion sowie Postleitzahl 2-, 3-, 4-, 5-Steller
gisd_score	Dezimalzahl	0 bis 1	Berechneter Score Sozioökonomischer Deprivation für die jeweilige Raumheinheit
gisd_5	Natürliche Zahl	1, 2, 3, 4, 5	Zugewiesenes GISD-Quintil der Raumeinheit gemäß der jährlichen Verteilung auf der jeweiligen räumlichen Ebene
gisd_10	Natürliche Zahl	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	Zugewiesenes GISD-Dezil der Raumeinheit gemäß der jährlichen Verteilung auf der jeweiligen räumlichen Ebene
gisd_k	Natürliche Zahl	1, 2, 3	1 unterstes Quintil, 2 drei mittlere Qunitile, 3 oberstes Qunitil
gemeinde_name, gvb_name, kreis_name, nuts2_name, ror_name	String	Flensburg, Stadt, Kiel, Landeshauptstadt, ..., Windischleuba	Name der Raumeinheit
Jahr	Natürliche Zahl	1998 bis 2019	Beobachtungsjahr

Formatierung der Daten

Die Daten sind im Datensatz als kommaseparierte `.csv`-Datei enthalten. Der verwendete Zeichensatz der `.csv`-Datei ist UTF-8. Trennzeichen der einzelnen Werte ist ein Komma ",".

- Zeichensatz: UTF-8
- Datumsformat: ISO8601
- `.csv`-Trennzeichen: ,

Zusätzlich wurden die Daten im Format `.dta` des Softwarepakets Stata v17 bereitgestellt.

Kontextmaterialien

Zusätzlich werden die Rohdaten aus der INKAR-Datenbank und die Syntax zur Generierung des GISD als Kontextmaterialien zur Verfügung gestellt.

[Kontextmaterialien/Rohdaten](#)
[Kontextmaterialien/Syntax/](#)

Im Repositorium werden außerdem frühere Versionen (Releases) für Replikationszwecke zum Download angeboten.

[Archiv](#)

Metadaten

Zur Erhöhung der Auffindbarkeit sind die bereitgestellten Daten mit Metadaten beschrieben. Über GitHub Actions werden Metadaten an die entsprechenden Plattformen verteilt. Für jede Plattform existiert eine spezifische Metadatendatei, diese sind im Metadaten-Ordner hinterlegt:

[Metadaten/](#)

Versionierung und DOI-Vergabe erfolgt über [Zenodo.org](https://zenodo.org). Die für den Import in Zenodo bereitgestellten Metadaten sind in der [zenodo.json](#) hinterlegt. Die Dokumentation der einzelnen Metadatenvariablen ist unter <https://developers.zenodo.org/#representation> nachlesbar.

[Metadaten/zenodo.json](#)

In der `zenodo.json` ist neben der Publikationsdatum (`"publication_date"`) auch der Datenstand enthalten:

```
"dates": [  
  {  
    "start": "2023-09-11T15:00:21+02:00",  
    "end": "2023-09-11T15:00:21+02:00",  
    "type": "Collected",  
    "description": "Date when the Dataset was created"  
  }  
],
```

Updates und Bugfixes

Ergänzung in der Version vom 26.5.2023:

- Dem Archiv wurde die Version GISD_Release_2022_v0.1 hinzugefügt

Ergänzung & Bugfix in der Version vom 03.01.2024:

- Dem Archiv wurde die Version GISD_Release_2022_v1.1 hinzugefügt
 - Bei der Generierung der GISD-Scores für die Postleitzahlebene blieben Postleitzahlen aus dem Kreis Göttingen unberücksichtigt. Die fehlenden Werte wurden eingefügt.
- Die [Syntax](#) zur Replikation der GISD Generierung wurde überarbeitet
- Die [Rohdaten](#) sind nun vollständig verfügbar. In den früheren Versionen waren die Zeitreihen der Rohdaten für die Indikatoren Beschäftigte ohne Abschluss und Beschäftigte mit akademischem Abschluss wegen unklarer Berechtigung der Weitergabe der Daten nicht vollständig.

Hinweise zur Nachnutzung der Daten

Offene Forschungsdaten des RKI werden auf [Zenodo.org](https://zenodo.org), [GitHub.com](https://github.com), [OpenCoDE](https://gitlab.opencode.de) und [Edoc.rki.de](https://edoc.rki.de) bereitgestellt:

- <https://zenodo.org/communities/robertkochinstitut>
- <https://github.com/robert-koch-institut>
- <https://gitlab.opencode.de/robert-koch-institut>
- <https://edoc.rki.de/>

Lizenz

Der Datensatz "SARS-CoV-2-Sequenzdaten aus Deutschland" ist lizenziert unter der [Creative Commons Namensnennung 4.0 International Public License | CC-BY](#) .

Die im Datensatz bereitgestellten Daten sind, unter Bedingung der Namensnennung des Robert Koch-Instituts als Quelle, frei verfügbar. Das bedeutet, jede Person hat das Recht die Daten zu verarbeiten und zu verändern, Derivate des Datensatzes zu erstellen und sie für kommerzielle und nicht kommerzielle Zwecke zu nutzen. Weitere Informationen zur Lizenz finden sich in der [LICENSE](#) bzw. [LIZENZ](#) Datei des Datensatzes..