Synthetic Data for Official Statistics

Simon Xi Ning Kolb, Destatis

Q 2024

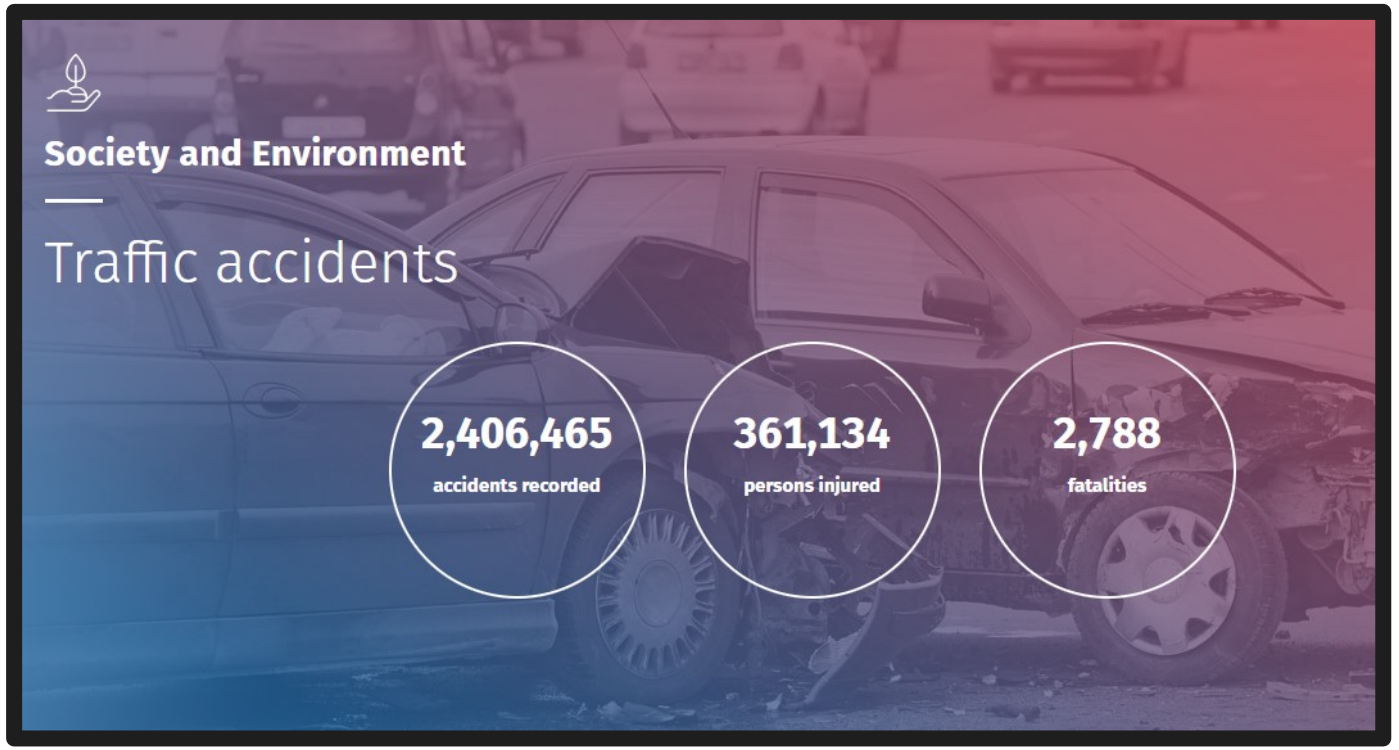
AnigeD

SPONSORED BY THE

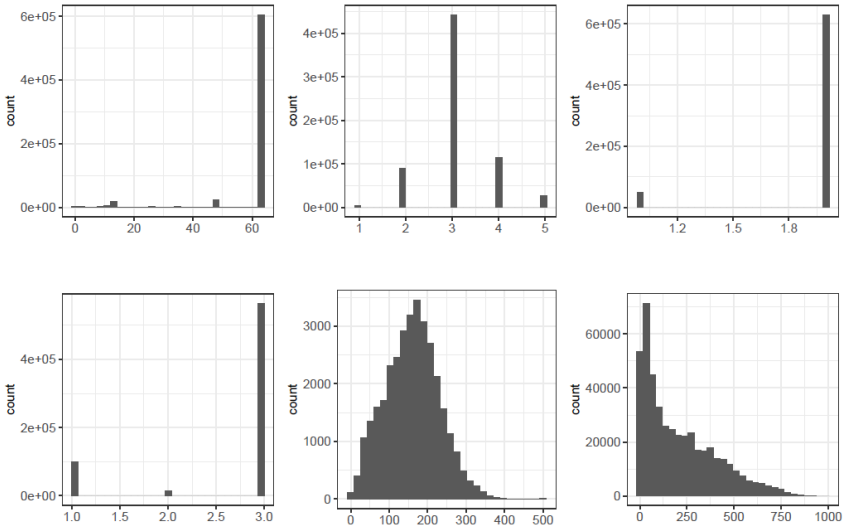
Federal Ministry of Education and Research

Funded by the European Union

NextGenerationEU


Society and Environment
Traffic accidents
2,406,465
accidents recorded
361,134
persons injured
2,788
fatalities

Microdata



N = 680.000

What data does **DESTATIS** publish?

Country level tables

Accidents and casualties in road traffic

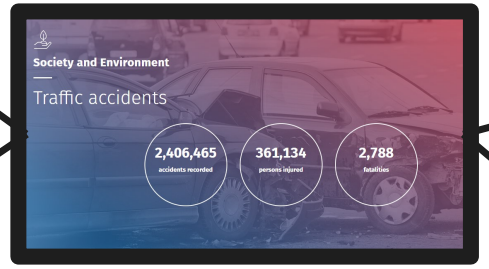
Specification	Unit	2023[1]	2022	2021	2020
Total number of accidents registered by the police	Number	2,512,899	2,406,465	2,314,938	2,245,245
of which					
Accidents involving personal injury	Number	289,726	289,672	258,987	264,499
Serious accidents involving material damage					
in the strict sense	Number	64,596	62,726	62,398	58,014
under the influence of intoxicating substances[2]	Number	15,453	15,649	13,815	13,596
Other accidents involving material damage	Number	2,143,124	2,038,418	1,979,738	1,909,136
Casualties, total	Number	366,478	363,922	325,691	330,269
of which					
Persons killed	Number	2,817	2,788	2,562	2,719
Persons seriously injured	Number	52,465	57,727	55,137	58,005
Persons slightly injured	Number	311,196	303,407	267,992	269,545

Accidents registered by the police

- Accidents registered by the police by type of damage/location
- Driver-related causes of accidents
- Accidents registered by the police: specification
- Long term series: Road traffic accidents

Persons involved and casualties

- Persons involved in accidents causing personal injury by type of traffic participation
- Main accident perpetrator
- Persons killed in road traffic accidents by month
- Persons killed in traffic accidents
- Persons killed in traffic accidents by type of traffic participation
- Persons injured in traffic accidents
- Persons injured in traffic accidents by type of traffic participation
- Hauptverursacher von Unfällen mit Personenschaden nach Segmenten und Unfallfolgen



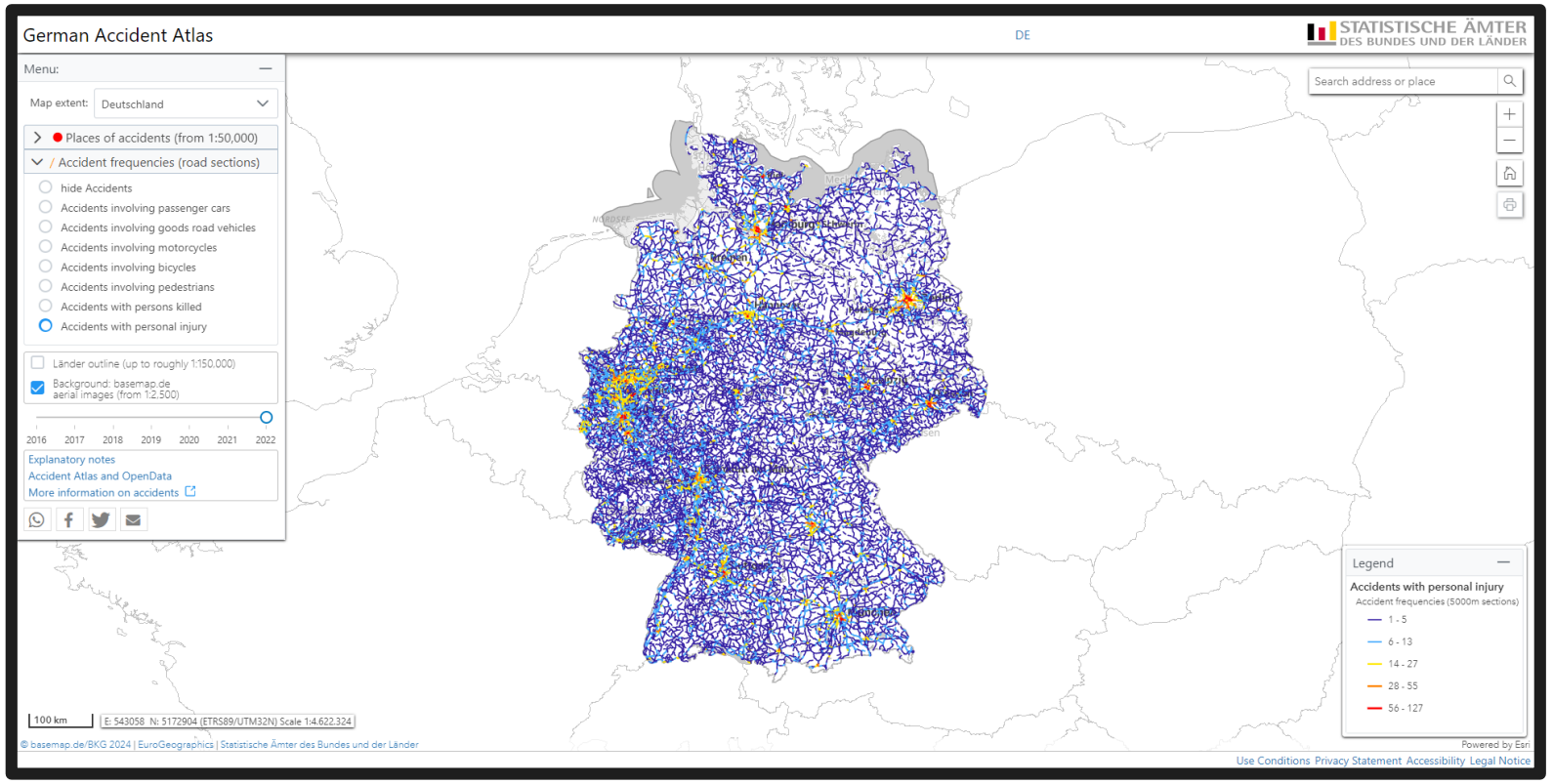
Causes of accident

- Driver-related causes of accidents
- Technical faults
- Road surface conditions, influence of the weather, obstacles
- Improper behaviour of pedestrians

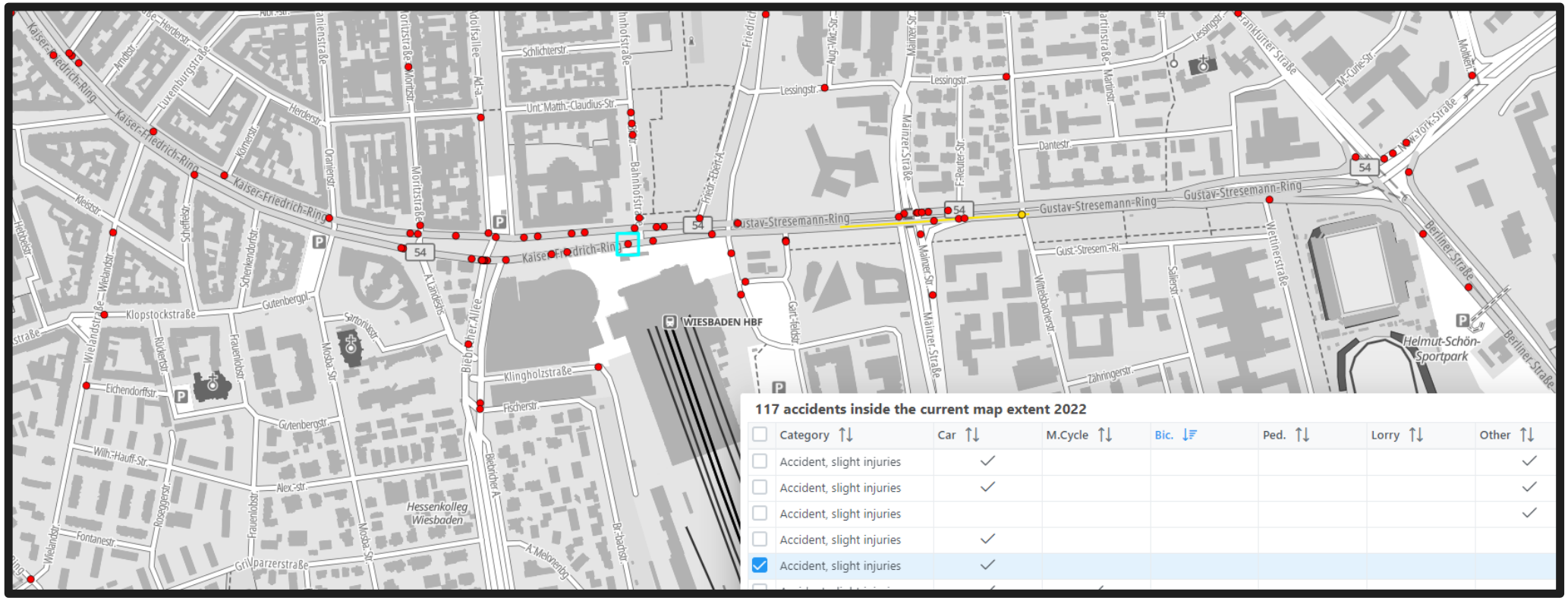
Rail traffic accidents

- Accidents and casualties in railway transport by group of persons
- Accidents and casualties in tramway transport by group of persons

Open data file – interactive accident atlas



Open data file - interactive accident atlas



What data do the **Offices of the Federal States** publish?

Publication program of the Federal States

- Frequency tables of varying granularity (state – district – municipality)
- *Many* two to five table dimensions (sparse tables)
- Additionally: unknown custom evaluations

Problems for conventional SDC (Cell Suppression/CKM)

State Tables	Open Data File
• Published margins	• Published margins

Partial synthesis

Participant	V1	V2	S1	S2
1	…	…	…	…
2	…	…	…	…
3	…	…	…	…
4	…	…	…	…
5	…	…	…	…
6	…	…	…	…
7	…	…	…	…
8	…	…	…	…
9	…	…	…	…

Non sensitive **sensitive**

Partial synthesis

Participant	V1	V2	S1	S2
1	…	…	…	…
2	…	…	…	…
3	…	…	…	…
4	…	…	…	…
5	…	…	…	…
6	…	…	…	…
7	…	…	…	…
8	…	…	…	…
9	…	…	…	…

Non sensitive **sensitive**

Evaluation on tabular level

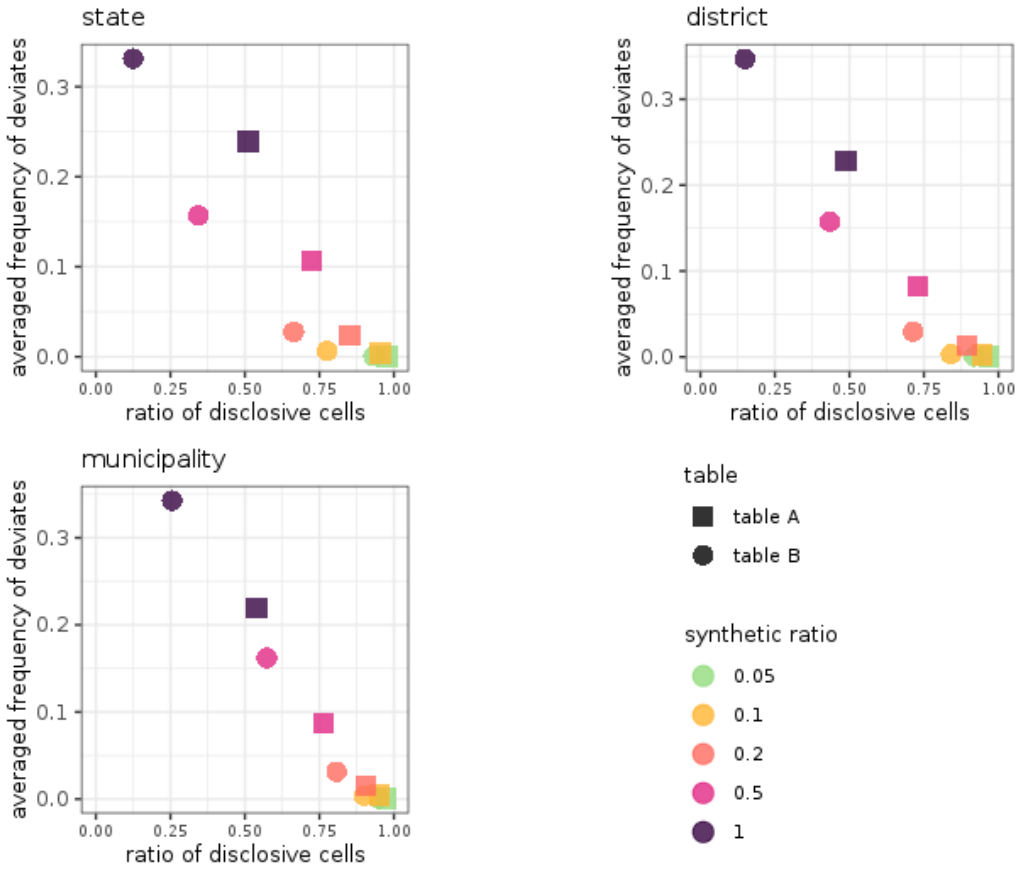
Utility

$$u(s,b) = \frac{1}{2}\left(\frac{n_{small}(s)}{N_{small}} + \frac{n_{big}(b)}{N_{big}}\right)$$

Risk

$$risk = \frac{common\ disclosive\ cells}{synthetic\ disclosive\ cells}$$

Risk-Utility Map for two tables



Contact

Statistisches Bundesamt
65180 Wiesbaden

Simon Xi Ning Kolb

www.destatis.de

www.destatis.de/kontakt

Email:
simonxining.kolb@destatis.de

Appendix

Risk

Original	L1	L2	L3
U1	11	0	0
U2	0	9	0
U3	2	0	0
U4	4	0	6

Risk

Original	L1	L2	L3
U1	11	0	0
U2	0	9	0
U3	2	0	0
U4	4	0	6

Synthesis	L1	L2	L3
U1	12	0	0
U2	0	9	0
U3	0	1	0
U4	0	0	7

Risk

Original	L1	L2	L3
U1	11	0	0
U2	0	9	0
U3	2	0	0
U4	4	0	6

Synthesis	L1	L2	L3
U1	12	0	0
U2	0	9	0
U3	0	1	0
U4	0	0	7

Messung Aufdeckungsrisiko

O	L1	L2	L3
U1	11	0	0
U2	0	9	0
U3	2	0	0
U4	4	0	6

S	L1	L2	L3
U1	12	0	0
U2	0	9	0
U3	0	1	0
U4	0	0	7

$$\frac{common\ disclosive\ cells}{synthetic\ disclosive\ cells} =$$

Messung Aufdeckungsrisiko

O	L1	L2	L3
U1	11	0	0
U2	0	9	0
U3	2	0	0
U4	4	0	6

S	L1	L2	L3
U1	12	0	0
U2	0	9	0
U3	0	1	0
U4	0	0	7

$$\frac{common\ disclosive\ cells}{synthetic\ disclosive\ cells} = \frac{(U1, L1) + (U4, L3)}{(U4, L3)} = \frac{1}{2} = 50\%$$