

# The Most Dangerous Districts of Dortmund

Tim Beige, Thomas Terhorst, Claus Weihs, and Holger Wormer

**Abstract** In this paper the districts of Dortmund, a big German city, are ranked concerning their level of risk to be involved in an offence. In order to measure this risk the offences reported by police press reports in the year 2011 (Presseportal, <http://www.presseportal.de/polizeipresse/pm/4971/polizei-dortmund?start=0>, 2011) were analyzed and weighted by their maximum penalty corresponding to the German criminal code. The resulting danger index was used to rank the districts. Moreover, the socio-demographic influences on the different offences are studied. The most probable influences appear to be traffic density (Sierau, Dortmundenerinnen und Dortmundener unterwegs—Ergebnisse einer Befragung von Dortmundener Haushalten zu Mobilität und Mobilitätsverhalten, Ergebnisbericht, Dortmund-Agentur/Graphischer Betrieb Dortmund 09/2006, 2006) and the share of older people. Also, the inner city parts appear to be much more dangerous than the outskirts of the city of Dortmund. However, can these results be trusted? Following the press office of Dortmund's police, offences might not be uniformly reported by the districts to the office and small offences like pick-pocketing are never reported in police press reports. Therefore, this case could also be an example how an unsystematic press policy may cause an unintended bias in the public perception and media awareness.

---

T. Beige (✉) · C. Weihs  
Chair of Computational Statistics, TU Dortmund, Germany  
e-mail: [tim.beige@gmx.de](mailto:tim.beige@gmx.de); [claus.weihs@T-Online.de](mailto:claus.weihs@T-Online.de)

T. Terhorst · H. Wormer  
Institute of Journalism, TU Dortmund, Germany  
e-mail: [ThomasTerhorst@gmx.de](mailto:ThomasTerhorst@gmx.de); [holger.wormer@udo.edu](mailto:holger.wormer@udo.edu)

## 1 Introduction

The paper is the result of a project together with data journalists, who were particularly interested in the prejudice of some citizens of Dortmund that the district Innenstadt Nord is especially dangerous. To find an answer to this question some methodological problems have to be overcome such as how danger can actually be measured. Section 2 gives information about the used response variables. In Sect. 3 different influential factors on danger in a district are discussed as well as outcomes of regression models. Finally, the development and analysis of a danger index is presented in Sect. 4.

## 2 Responses

How can danger be measured? This project deals with different types of offences. It is based on a data set which is created by the Dortmund police press releases of the year 2011 (see [www.presseportal.de](http://www.presseportal.de)). In our case, incidents or non-crime reports such as demonstrations, announcements, public relations and similar press reports are not considered. Overall, 1,053 press releases are cataloged. Each data row or rather each offence contains several variables. The variable *offence* specifies the type of offence. This nominally scaled variable includes the considered characteristic attributes of this project, which are recorded in Table 3 among other aspects. Figure 1 shows a descriptive analysis of the frequency of offences which are used as response variables  $y$  later in Sect. 3.3, see also Table 3. Furthermore, there is a nominally scaled variable specifying *the district* in which the offence took place. These are the districts of Dortmund as defined by official statistics. In addition to these variables, *the street* is given for a more accurate description of the crime scene. If a street is located in several districts it has been allocated to that district in which the longer part of the street is situated.

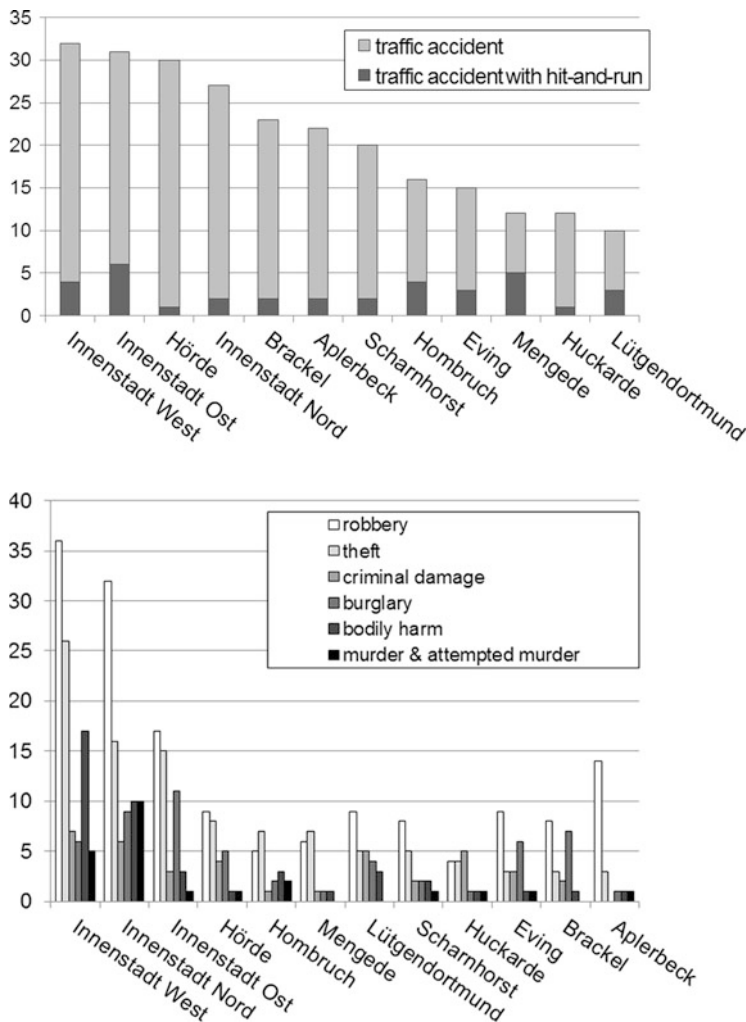
There are deficits in the response data set. Some of the variables have missing values, others are inaccurate. Only the complete 825 press releases are used for counting the offences in the 12 districts of the city of Dortmund.

## 3 Modeling the Offences

This section describes the statistical modeling of offences.

### 3.1 Factors of Influence

Let us first discuss different factors  $x_i$  of influence. First, socio-demographic factors are considered such as *population density* given in inhabitants per hectare,



**Fig. 1** Types of offences used in the study and their occurrence in the districts

the registered *number of inhabitants* in the district, the number of recorded *unemployed people* in the district, the *unemployment rate*, i.e. the unemployed people as a share of labor force (15–65 years old people), as well as the number of *social security recipients* (persons getting SGB-II). Furthermore, the *share of migrants* in the population is considered as a characteristic of the social structure of a district.

In addition, as a geographic factor the *area size* given in hectare is available in the data set. The demographic factors examined in the project are all ratios to labor force, including the *youth rate* for under 15 year old people, the *old rate* for over 65 year old people, as well as the *very old rate* for over 80 year old people. Moreover,

**Table 1** Number of ways between starting point and destination

Table: starting points and destinations of all ways															
from ... \ to ...	In-West	In-Nord	In-Ost	Eving	Scharnhorst	Brackel	Aplerbeck	Hörde	Hornbruch	Lütgendortmund	Huckarde	Mengede	Dortmund	Outside	Overall
In-West	1,014	168	163	53	65	123	115	126	170	126	92	62	2,277	149	2,426
In-Nord	162	646	68	39	22	27	19	40	21	17	21	18	1,100	73	1,173
In-Ost	190	73	548	11	8	68	43	72	63	19	15	7	1,117	81	1,198
Eving	63	49	9	398	28	17	13	7	8	2	8	10	612	56	668
Scharnhorst	80	25	13	33	492	88	26	9	14	7	11	5	803	75	878
Brackel	146	23	66	16	73	761	60	39	24	17	9	8	1,242	85	1,327
Aplerbeck	128	22	40	12	21	65	707	91	26	29	14	6	1,161	122	1,283
Hörde	143	48	71	9	8	44	81	618	130	38	18	8	1,216	126	1,342
Hornbruch	191	18	44	7	15	27	33	122	689	57	14	6	1,223	122	1,345
Lütgendortmund	142	20	17	3	7	20	21	35	53	717	103	20	1,158	169	1,327
Huckarde	102	21	23	10	15	9	5	18	12	89	440	30	774	60	834
Mengede	76	16	9	13	4	7	7	10	5	19	35	406	607	69	676
Total	2,437	1,129	1,071	604	758	1,256	1,130	1,187	1,215	1,137	780	586	13,290	1,187	14,477
Dortmund	2,438	1,136	1,071	607	760	1,257	1,140	1,196	1,221	1,144	781	590	13,341	1,205	14,546
Outside	97	47	58	50	54	58	104	116	108	144	49	43	928	75	1,003
Overall	2,535	1,183	1,129	657	814	1,315	1,244	1,312	1,329	1,288	830	633	14,269	1,280	15,549

some political factors of influence are examined, too, namely the shares of extreme *right-wing* and *left-wing* voters in the local elections of 2009 . Here, extreme left means voters of the party “Die Linke”, while extreme right means voters of the party DVU.

Finally, there is an indicator for *traffic density* based on a representative survey study in 2005 (Sierau 2006). Table 1 contains the number of ways passed inside Dortmund on the representative day of the survey. It includes ways inside a district and ways from one district to another. For a district, not only the number of ways inside a district is counted, but also all outgoing and incoming ways to the district, as well as all intersecting ways going through the district when starting point and destination are not in adjacent districts. These intersecting ways are determined by using the main streets of Dortmund, precisely the A40, A45, B1, B54, B236 and the Malinckrodtstraße are chosen as main traffic routes. After all ways had been considered, an index is obtained which can be used as a measure of traffic density.

### 3.2 Quality of Influence Data

The factors of influence in the data set are not of identical quality. All variables except of the traffic density and the election results are taken from the annual statistical report of Dortmund from 2011. Similarly accurate are the election results which have been determined by the city of Dortmund. They are based solely on elections in 2009. The traffic density relies on a study in 2005 (see Sierau 2006, p. 32) and is estimated using the method described in Sect. 3.1. This variable may not be that accurate, particularly since the estimates are based solely on the motion profile of Dortmund citizens. Any commuters or tourists are not considered.

**Table 2** Correlations between possible factors of influence

	Ways	UR	Pop d	Migr.	Yr	Or	Right	Left
Traffic density	1.00							
Unemployment rate	−0.07	1.00						
Population density	0.77	0.31	1.00					
Share of migrants	−0.05	0.95	0.33	1.00				
Youth rate	−0.80	0.52	−0.48	0.54	1.00			
Old rate	−0.23	−0.85	−0.50	−0.82	−0.20	1.00		
Extreme right-wing voters	−0.47	0.28	−0.32	0.19	0.43	−0.35	1.00	
Extreme left-wing voters	0.10	0.95	0.44	0.95	0.40	−0.86	0.14	1.00

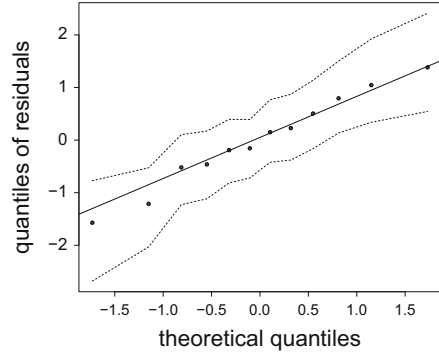
### 3.3 Regression Results for the Factors of Influence

For statistical modeling we use a linear model of the form  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$ , where the response  $y$  and the influential factors  $x_i$  are observed for the 12 districts of the city of Dortmund, and  $\varepsilon$  is normally distributed with zero mean and variance  $\sigma^2$ . Model fitting is based on the AIC criterion with backward selection. For ease of interpretation, it is helpful that the variables included in the model are (at least nearly) uncorrelated. Therefore, the Bravais–Pearson correlations of the influential factors are examined first. The variable *social security recipients* is particularly highly correlated with *unemployed people* and *unemployment rate*, and also *population density* with *area* as well as the *very old rate* with the *old rate*. It is obvious that one of these variables may (at least nearly) sufficiently explain the corresponding other variables. Therefore it is merely sufficient to put just one of these variables in the model. Also the variable *population* is not included in the model, but only the corresponding rate “population density” in order to avoid size effects in model building. Based on this, we decided to only consider eight rates in the model, namely *traffic density*, *unemployment rate*, *population density*, *share of migrants*, *youth rate*, *old rate*, and the shares of the right-wing and left-wing voters. Please note, however, that due to an outlier there are still some high correlations between some influential factors (see Table 2).

Using these eight factors in the linear model let us first check, whether the normality assumption of the linear model is fulfilled. Figure 2 shows a QQ-plot illustrating that the model residuals are quite normal, e.g., for the response variable criminal damage. Also for the other offences, the assumption is reasonably fulfilled.

As examples, let us roughly discuss the greatest influences on traffic accidents and robberies in the models. A more detailed model discussion will be given for the danger index in Sect. 4. Not surprisingly, for traffic accidents this is traffic density (p-value in t-test equal to 0.0068). Higher traffic density leads to more accidents. But also a higher old rate increases the risk of accidents. Even on robberies the indicator for traffic appears to be significantly positive. That can be explained by the fact that when there are more people the probability of an attack is higher. On the contrary, a high old rate reduces risk of robberies. In general, the traffic density is usually

**Fig. 2** QQ-plot for the model with the response variable criminal damage with an estimated confidence interval at the 95 % level for each data point



positively significant because it is an indicator for the number of people which spend time in the district. The goodness of fit of the models was always high. Having realized one outlier in the influential factors,  $L_1$ -methods could be an alternative for model building. Fortunately, elimination of the outlier did not change the regression coefficients too much. For model building the software R was used ([R Development Core Team 2011](#)).

## 4 Danger Index

Finally, a danger index is created whose value indicates the overall danger for each district. In order to weight the different types of crime, the German criminal code is used. Each offence obtains the maximal penalty as a weight (see [Table 3](#)). After that, the weighted offences are summed up as an index for each district. [Table 4](#) shows the values of the calculated danger index for the districts of Dortmund while [Fig. 3](#) illustrates the results using a grey-shaded city map. According to this index, Innenstadt Nord has the highest danger potential while Huckarde would be the safest district in Dortmund. Let us examine which influential factors affect the overall danger. The variables selected by backward selection with AIC are shown in [Table 5](#) together with their estimated coefficients  $\hat{\beta}$  and p-values. Moreover, we report the corresponding values  $\hat{\beta}_s$  for the model without the insignificant factors. This elimination only leads to a small decrease of  $R^2$  from 0.952 to 0.944. Our modeling shows that a high share of extreme right-wing voters as well as a high old rate is associated with lower danger in a district. Contrary to this, the traffic density significantly increases the risk of becoming a victim of crime or accidents.

**Table 3** Offences with their maximum penalty in years

Offence	Max penalty	Offence	Max penalty
Traffic accident with hit-and-run	3	Bodily harm	5
Dangerous disruption of air traffic	10	Attempted bodily harm	5
Traffic accident	0 (0.1) <sup>a</sup>	Criminal damage	2
Wrong way driver	5	Suicide	0
Robbery	5	Attempted suicide	0
Aggravated robbery	10	Murder	$\infty$ (25) <sup>b</sup>
Attempted robbery	5	Attempted murder	$\infty$ (25) <sup>b</sup>
Burglary	10 (5) <sup>c</sup>	Missing	0
Theft	5	Breach of public peace	3
Attempted theft	5	Drug possession	5
Hit-and-run	3	Animal murder	3
Receiving stolen goods	5		

<sup>a</sup> In the German criminal code (StGB) no imprisonment for traffic accidents is provided. To add this delict nevertheless, it is evaluated with the weak factor of 0.1

<sup>b</sup> In accordance to sec. 211 §(1) murder is in Germany liable by life imprisonment. This penalty means to be liable for an indefinite period, but at least 15 years. In addition to this, there are 5 years probation because of sec. 56a §(1), plus 2 years suspension period of eligibility of the convict in according to sec. 57a §(4). Therefore, the total penalty for murder is 22 years. Afterwards the convict can put into an application for releasing from custody. Currently, these applications are usually accepted after the second or third application, therefore after 2 or 4 years. For calculating the danger index, murder is thus indicated by  $22+3=25$

<sup>c</sup> According to StGB sec. 243 §(1) No. 1 burglary can be penalized with up to 10 years in prison. This occurs only in exceptional cases. In general, burglary is a case of section 243 §(1) No. 2 and is listed with 5 years imprisonment as maximum penalty

**Table 4** Danger index for 2011

District	Danger index	District	Danger index
Innenstadt Nord	666.50	Lütgendortmund	134.70
Innenstadt West	635.80	Hörde	133.90
Innenstadt Ost	318.50	Scharnhorst	126.80
Hombruch	160.20	Brackel	123.10
Eving	156.20	Mengede	112.70
Aplerbeck	136.00	Huckarde	89.10

## 5 Problems with the Data Base

The Dortmund police press office who was finally confronted with the results of our estimation warns about over-interpreting the study. They stress that one should have in mind that the published press releases are chosen solely for media. This means it is decided how interesting the publications of the incidents are, e.g., for the newspapers in Dortmund. Smaller offences like pick-pocketing are barely mentioned. Therefore, the evaluation of the releases could draw a wrong picture of the offence distribution. Also, according to the press office some colleagues on the

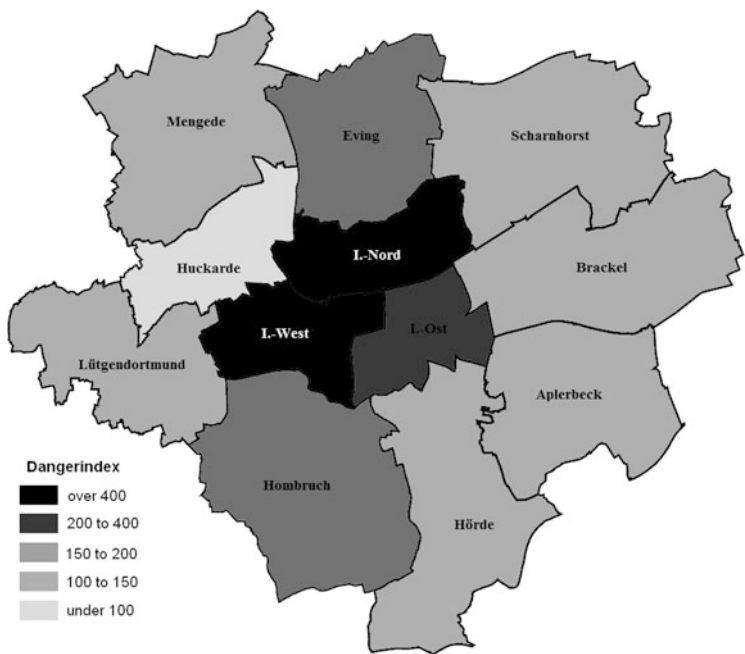


Fig. 3 Danger index in the city of Dortmund

Table 5 Regression results for the danger index after backward selection

Factor of influence	$\hat{\beta}$	$ t $	$p$ -value	$\hat{\beta}_s$	$ t_s $	$p$ -value <sub>s</sub>
Intercept	1,000.867	2.676	0.0440	1,241.062	8.932	< 0.0001
Population density	−2.382	−0.997	0.3646			
Traffic density	0.085	2.844	0.0361	0.051	2.784	0.0238
Unemployment rate	−11.001	−1.042	0.3452			
Share of migrants	7.160	1.878	0.1192			
Old rate	−25.585	−3.647	0.0148	−30.062	−10.230	< 0.0001
Share of right-wing	−72.994	−3.207	0.0238	−87.738	−4.465	0.0021

various police stations have to be encouraged to submit their press releases while others lavish with releases on the press office. This effect may be reflected in the ranking.

6 Summary and Outlook

This study examined the danger in the districts of Dortmund in the year 2011. Offences are weighted by means of the corresponding maximum penalty in the German criminal code. Different regression models estimate whether some influential



factors increase or decrease the risk of an offence. The general conclusion might be that more crimes occur where lots of people meet together, so when the traffic density is high. In contrast to this, districts with higher old rates reduce the risk of robberies, thefts and bodily harm, but increase the risk of accidents significantly.

A major problem within this project is the small sample size. Although sufficient offences (1,053) are identified, which are spread on just 12 districts. Because of this, it is difficult to analyze the factors of influence properly. In particular, the sensitivity with respect to outliers is large. It would be advisable to extend the study to North-Rhine-Westphalia or to use the statistical sub-districts of Dortmund to enlarge the sample size. In the latter case it would be necessary to observe the offences for a longer time period to obtain a sufficient number of observations. Perhaps this also could decrease the correlation problem. Furthermore, other influential variables could be examined, e.g. social welfare data or an indicator which classifies a shopping district. This could clarify the results overall.

## References

- R Development Core Team. (2011). *A language and environment for statistical computing*. Wien, Österreich: R Foundation for Statistical Computing. URL <http://www.R-project.org/>. ISBN 3-900051-07-0.
- Sierau, U. (2006). *Dortmunderinnen und Dortmunder unterwegs - Ergebnisse einer Befragung von Dortmunder Haushalten zu Mobilität und Mobilitätsverhalten, Ergebnisbericht*. Dortmund-Agentur/Graphischer Betrieb Dortmund 09/2006.

Data Analysis, Machine Learning and Knowledge  
Discovery

Spiliopoulou, M.; Schmidt-Thieme, B.; Janning, R. (Eds.)

2014, XXI, 470 p. 120 illus., 32 illus. in color., Softcover

ISBN: 978-3-319-01594-1