Research Proposal: Social Capital and Crime in Germany

Daniel Salgado Moreno and Lars Mehwald

13 November 2015

Introduction and theory

Most of crime studies are based on the rational choice assumptions of legal and criminal returns and the importance of deterrence to reduce criminal behaviours (Becker 1968). However, less has been written about the relevance of social interactions explaining criminal behaviours. Some literature, though, has focused on the relationship between violent crimes and social capital. Lederman (Lederman, Loayza, and Menedez 2002) identified that the prevalence of trust on community members has an effect on violent crimes and that the social cohesion through participation and membership in community organisations only impact violent crime levels depending on the type of religious and social organisation engaged. Additionally, Hanslaimer (Hanslmaier 2015) showed that altruism has a general negative effect on violent crimes, but other indicators of social capital only proved to be related to other types of high impact crime and not to violent crime. Hence, we are interested in responding to the question: How do social networks and bonds in a given community determine the level of violent crime?

Social capital can be summerised as the links, shared values, and understandings in society that enable individuals and groups to trust each other and so work as a group (OECD, n.d.). The concept of social capital gained importance after Robert Putnam used it to describe the anomic state of the American society at the beginning of the 21st century (Putnam 2001). He argued that a decrease in community linkages that once held societies together, translates in a loss of social capital (Putnam 2001).

Additionally, we expect to observe that in societies with anomic and disorganization symptoms there are lower levels of social capital. Within these societies, groups and member interactions find themselves in a disruptive phase, damaging social bonds, norms, values, understandings and commitment. Environments where people lack a strong moral order lead to individual members to behave more egoistically therefore diminishing social capital. Disintegration of members of the community weakens social cohesion and promotes individualism that end in 'Machiavellian' competition, fostering feelings of status insecurity. Moreover, certain social milieus tend to be more vulnerable and respond therefore with violence. In addition, social capital provides means for informal punishment within a given community. Therefore more social capital leads to more abiding to social norms. In contrast communities that lack of social capital can merely punish norm diverging behaviour informally to a lesser extent and therefore are expected to have a higher crime rate.

Our main hypothesis is that the more social capital a community has, the less crime there would be observed as a result of its collective action mechanisms. Thus, we expect to observe that communities with anomic symptoms will tend to have more crime.

Social Capital is a multi-dimensional concept that consist of "varios social elements that promote individual and collective action"(Lederman, Loayza, and Menedez 2002). As a consequence of its high complexity, measuring social capital carries some difficulties. However, researchers have shown consistent results when using indicators for social capital based on participation in community organisations, attitudes of civil action, and the sense of membership to the community (Lederman, Loayza, and Menedez 2002). Even more complex is the relation between social capital and crime, since the causal connection between the two may go in both directions, meaning that crime also can impact social capital. More violent crime could disrupt community bonds and reduce the social capital, or it could encourage some form of social capital through the formation of community engagement in fighting crime.

With new data available for crime statistics for the years 2013 at the districts' level in Germany it is relevant to study the incidence of high impact crimes and its relation to social capital formation. Results might be meaningful as rational for public intervention using crime reduction strategies that help build-up social capital at the community level (e.g. funding of sport clubs, gathering community, subsidies for staying in a community), instead of more aggressive interventionist policing strategies. Strategies that are less police intensive might have higher public support and might be seen as more legitimate. Social engagement at the local level might also help increase trust levels across community members and encourage more altruistic behaviours among them. Furthermore, social capital investment might be more cost-efficient compared to more aggressive policing tactics, which effects tend to be perceived only in the more immediate time, whereas social capital investments and their effects might stretch out for longer time.

Data

Data sources:

We have data from three main sources: 1. German Federal Police's crime statistic ("Polizeiliche Kriminalstatistik" [PKS]) 2. The Regional Database Germany published by the Federal Statistical Office and the statistical Offices of the Länder. 3. German Federal Foundations Association ("Bundesverband Deutscher Stiftungen")

Dependent Variable

Crime statistics were obtained from the German Federal Police's crime statistic ("Polizeiliche Kriminalstatistik" PKS) for only one year 2013 on a district level ("Kreis"). The first limitation we encounter was that the PKS is not providing time series for district level crime for more years than 2013 and 2014. However, we were not able to use the most recent data for the year 2014, because socio economic data at the district level in the Regional Database Germany is mostly available only before 2013 and in some cases only before 2012 for all districts.

This data is also disaggregated regarding the type of crime. One limitation to this statistics, however, is that they are missing the dark figure of crime –the unreported number of crime in this area. Different methods exist to assess the dark figure of crime, but it is difficult to assess their reliability. Hence, depending on the type of crime, the reported numbers might contain a structural bias, i.e. some types of crime are reported at a different rate than others (e.g. when insurance companies require individual reporting). Since we focus on crimes that disrupt the social community, we expect a high level of reporting and this reporting to be consistently high across different types of disruptive crimes. It is possible to focus on one type of crime for our dependent variable, but also to construct a dependent variable summarising different types of crime. The latter approach might prove utile, since this provides greater variance in the dependent variable.

Following this approach, we construct two measures for violent and non-violent crimes, for which we used a simple addition method, including all crimes fitting each category. After this addition, we then proceed by creating crime rates for total, violent, and non-violent crimes. This computation will allow for better comparison between districts.

Furthermore, given the limitations of our data we have to assume that this cross-sectional data for district crime are a simple count of independent events. This is relevant to remember for the estimation process.

Independent Variables

All data is provided by the German statistical bureau. We want to measure different dimensions of our independent variable social capital. First, we used the data on foundations density in every district as an indicator of community engagement. Foundation density is an indicator of the total number of foundations in each German district per 100 thousand inhabitants.

A second indicator for social capital we used was the total flow of migrants coming in and out of each district. This indicator served as an approximation to the concept of community membership. The assumption behind this indicator is that the more people moving in and out of a district, the less a person leaving in that district feels part of the community, which is might be also true for people moving. Since social capital requires time to build up, we expect higher crime rates in areas with high fluctuations.

The third dimension of social capital we tried to measure was the civil action within a community. We then looked at the voting turnout during the federal elections of 2013 in Germany to estimate the effect of social capital on crime through the participation of community members in policy decisions while deciding on the next government.

Moreover we need district level data for other variables to control for local heterogeneity. We want to control for factors that influence the rational choice for committing a crime. First, unemployment changes the opportunity costs of punishment, i.e. imprisonment. In addition we want to control for a district's composition in age (younger persons are found to be more involved in crimes), gender (men are found to be more related to crimes).

Summary statistics for all variables

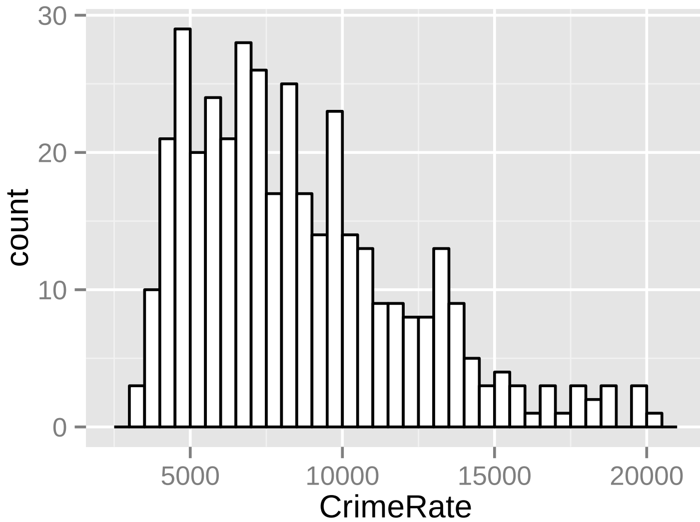
Table 1:

Statistic	N	Mean	St. Dev.	Min	Max
UnemployedPercentage	393	7.117	3.220	1.500	16.800
TurnoutPercentage	393	70.785	4.241	57.800	79.100
FoundationsDensity100k	393	21.162	14.933	2.700	89.200
MarriageRate	393	477.312	97.283	291.029	1,282.091
MaleRate	393	49.030	0.669	46.957	51.002
YouthRate	393	5.987	4.150	0.225	24.872
BelieversRate	393	99,816.400	1,729.588	94,622.830	103,798.100
FlowRate	393	12,105.840	2,909.114	6,427.000	26,410.970
ViolentCrimeRate	393	791.773	355.100	271.157	2,266.318
NonViolentCrimeRate	393	7,826.304	3,303.962	2,870.475	18,662.010
CrimeRate	393	8,618.077	3,604.079	3,141.632	20,007.840

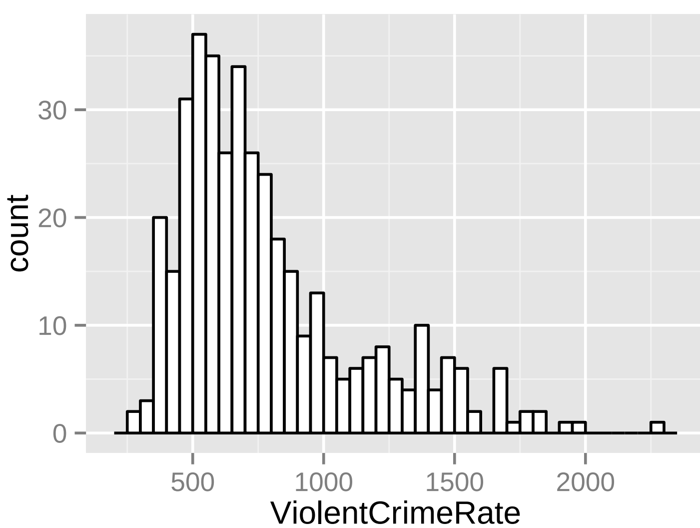
Methodology and analysis

For this research, we have conducted a cross-district study to estimate the relation of different indicators of social capital –engagement in the community, political participation and community stability– and the incidence of violent and non-violent crime. In order to answer the research question, we will apply a linear regression model in R Studio (R Core Team 2015). So far we have used R packages by the following authors: Wickham and Francois (2015), Wickham and Chang (2015), Gandrud (2015), C.-h. Chan et al. (2015) and Wickham (2015). A binomial or multinomial regression model does not qualify, since we have a continuous, interval level dependent variable. Since our data only covers two time periods (t=2) we intend to use pooled OLS regressions. We thereby might run the risk to violate OLS assumptions when the values for a district from one time period are correlated with the observation for this district at the other time period - assuming that the environment remains rather constant and districts with higher crime rates will also experience higher crime rates in the next period. Therefore we need to check statistically, whether our measured estimators are unbiased and efficient (e.g. comparing the results to a panel regression model, dummy for time period, heteroskedasticity).

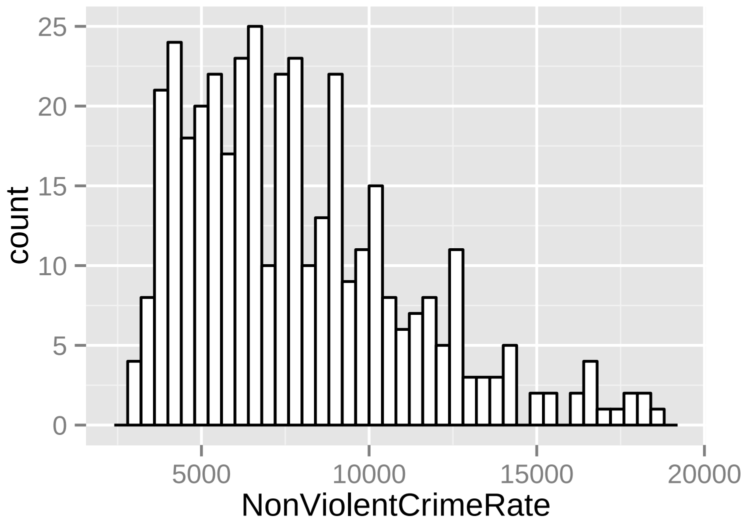
This histogram shows the frequency distribution of the independent variable crime rate (total number of crimes per 100,000 individuals) for all districts



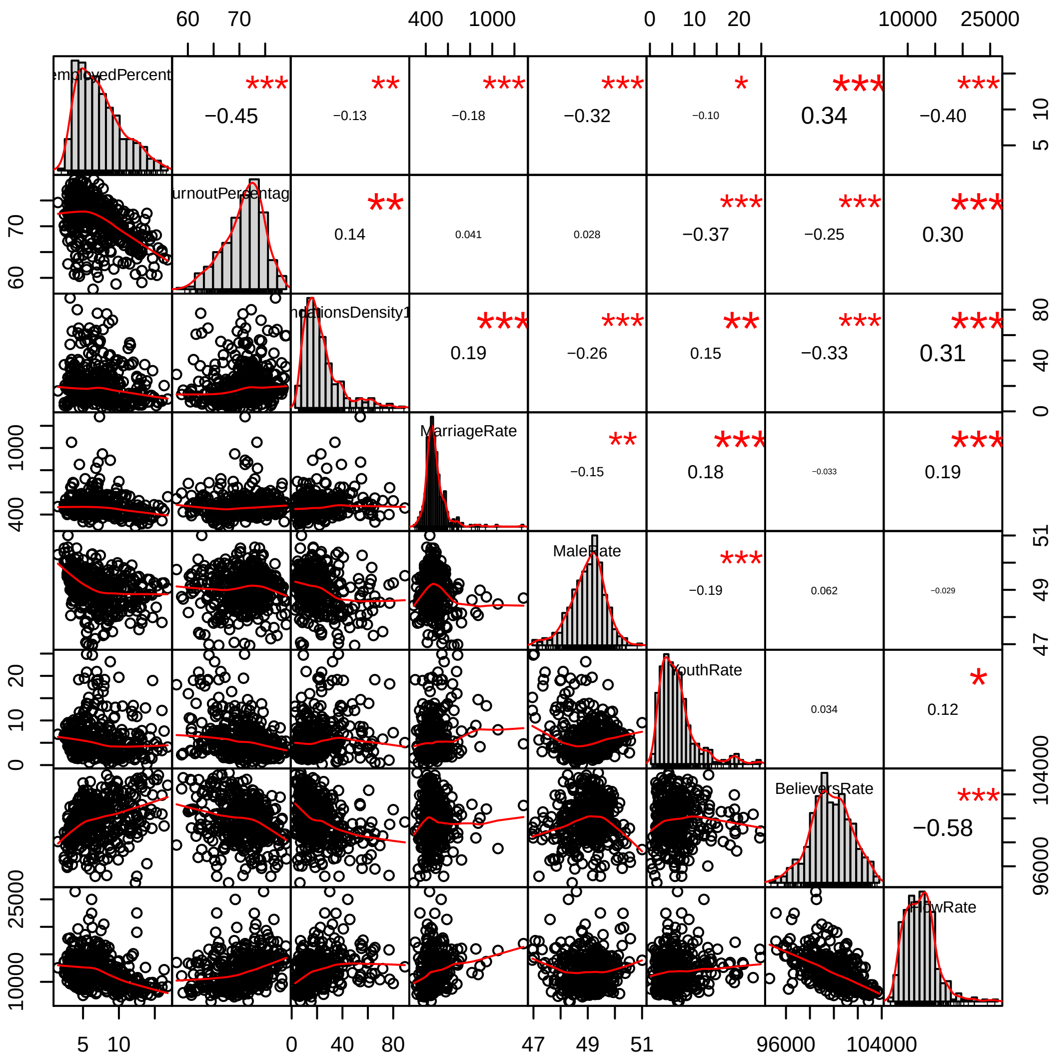
This histogram shows the frequency distribution of the independent variable violent crime rate (total number of violent crimes per 100,000 individuals) for all districts



This histogram shows the frequency distribution of the independent variable non-violent crime rate (total number of non-violent crimes per 100,000 individuals) for all districts



This correlation matrix shows the correlation of the independent and control variables. The correlation coefficient and their significance are shown in the upper right corner. The bivariate scatter plots are shown in the lower left corner. The histogram for all independent variables are shown on the diagonal.



Model specifications

Because crime events are discrete events, the possible crime rates for any given populations size are those corresponding to integer counts of crimes. Change in populations sizes will lead to violating the assumption of homogeneity of error variance. Second, normal or symmetrical error distributions of crime rates cannot be assumed. So the error distribution is skewed to the lower bound (to the left). Thus, a more formal way to accommodate over-dispersion in a count data regression model is to use a negative binomial model as tables 2 to 5 show.

Optimal data set

Throughout the analysis we face severe data constraints that inhibit an in depth analysis of the theoretical relationship. On optimal data set would contain the following. It would contain time series data for every unit of observation. This would enable us to use a fixed effects model (since we already have cross-sectional data) and thereby controlling for unobserved heterogeneity between the observations. This data would also enable us to address the most severe challenge to this analysis: reverse causality. People are deemed to respond to a criminal environment and it is likely that this response is correlated with our independent variables (i.e. people with higher social capital tend to move out of areas with high crime rates). On optimal dataset would also contain more detailed data. This would allow more elaborate controls and better measurement of social capital. In addition individual level data would improve the analysis by allowing for a multi level analysis controlling for individuals characteristics. This could also address the main challenge of reversed causality since migration could be explicitly modelled.

Conclusion

A preliminary analysis of the model specification shows that the negative binomial model does not change the statistical significance of the main explainatory variables. Moreover, only some independent variables attain statistical significance for violent crimes. Further model specifications are require in order to improve the reliability of these findings. However, further estimations are needed to improve the interpretations of the coeficients, since the estimated coefficents are log likelihoods.

Regression analysis

[1]

Table 2: Regression analysis regarding (non-) violent and total crimes and independent variable FoundationsDensity100k with OLS regressions and negative binominal regression

	Dependent variable:			
	ViolentCrimeRate	NonViolentCrimeRate	CrimeRate	
	OLS	*OLS*	*OLS*	*negative binomial*
	(1)	(2)	(3)	(4)
FoundationsDensity100k	3.80***	20.64***	24.43***	0.002***
	(0.85)	(6.30)	(6.92)	(0.001)
BelieversRate	−0.08***	−0.77***	−0.85***	−0.0001***
	(0.01)	(0.06)	(0.06)	(0.0000)
MarriageRate	−0.23*	−1.15	−1.39	0.0000
	(0.12)	(0.92)	(1.01)	(0.0001)
MaleRate	−84.03***	−884.99***	−969.02***	−0.10***
	(19.94)	(147.01)	(161.46)	(0.02)
YouthRate	29.16***	54.71**	83.87***	0.01***
	(2.90)	(21.38)	(23.48)	(0.003)
UnemployedPercentage	66.83***	824.91***	891.74***	0.10***
	(4.28)	(31.58)	(34.68)	(0.004)
Constant	11,789.58***	122,216.00***	134,005.60***	21.72***
	(1,159.26)	(8,548.98)	(9,388.82)	(1.04)
Observations	393	393	393	393
R^2	0.60	0.75	0.75	
Adjusted R^2	0.59	0.75	0.74	
Log Likelihood				−3,485.04
θ				21.82*** (1.55)
Akaike Inf. Crit.				6,984.09
Note:				*p<0.1; **p<0.05; ***p<0.01

This regression output shows the results using 3 different dependent variables and two models

Table 3: Regression analysis regarding (non-) violent and total crimes and independent variable FlowRate with OLS regressions and negative binominal regression

	Dependent variable:			
	ViolentCrimeRate	NonViolentCrimeRate	CrimeRate	
	OLS	*OLS*	*OLS*	*negative binomial*
	(1)	(2)	(3)	(4)
FlowRate	0.01**	0.08**	0.09**	0.0000***
	(0.01)	(0.04)	(0.04)	(0.0000)
BelieversRate	−0.08***	−0.75***	−0.83***	−0.0001***
	(0.01)	(0.06)	(0.07)	(0.0000)
MarriageRate	−0.20	−1.07	−1.27	−0.0000
	(0.13)	(0.93)	(1.02)	(0.0001)
MaleRate	−102.07***	−979.93***	−1,082.00***	−0.11***
	(19.84)	(144.62)	(159.12)	(0.02)
YouthRate	29.69***	56.16***	85.85***	0.01***
	(2.96)	(21.58)	(23.75)	(0.003)
UnemployedPercentage	67.44***	832.51***	899.95***	0.10***
	(4.46)	(32.52)	(35.78)	(0.004)
Constant	12,617.15***	124,374.70***	136,991.90***	21.01***
	(1,219.13)	(8,886.69)	(9,777.63)	(1.08)
Observations	393	393	393	393
R^2	0.59	0.75	0.74	
Adjusted R^2	0.58	0.74	0.74	
Log Likelihood				−3,481.86
θ				22.17*** (1.57)
Akaike Inf. Crit.				6,977.72
Note:				*p<0.1; **p<0.05; ***p<0.01

This regression output shows the results using 3 different dependent variables and two models

[1]

Table 4: Regression analysis regarding (non-) violent and total crimes and independent variable TurnoutPercentage with OLS regressions and negative binominal regression

	Dependent variable:			
	ViolentCrimeRate	NonViolentCrimeRate	CrimeRate	
	OLS	*OLS*	*OLS*	*negative binomial*
	(1)	(2)	(3)	(4)
TurnoutPercentage	-10.76^{***}	-12.54	-23.30	0.001
	(3.63)	(26.75)	(29.42)	(0.003)
BelieversRate	-0.09^{***}	-0.82^{***}	-0.91^{***}	-0.0001^{***}
	(0.01)	(0.05)	(0.06)	(0.0000)
MarriageRate	-0.18	-0.80	-0.98	0.0001
	(0.13)	(0.92)	(1.02)	(0.0001)
MaleRate	-120.64^{***}	$-1{,}010.89^{***}$	$-1{,}131.53^{***}$	-0.11^{***}
	(20.54)	(151.45)	(166.60)	(0.02)
YouthRate	25.18^{***}	55.25^{**}	80.43^{***}	0.01^{***}
	(3.41)	(25.15)	(27.66)	(0.003)
UnemployedPercentage	57.28^{***}	807.79^{***}	865.07^{***}	0.10^{***}
	(5.13)	(37.79)	(41.57)	(0.005)
Constant	$15{,}499.91^{***}$	$134{,}840.40^{***}$	$150{,}340.30^{***}$	22.73^{***}
	(1,240.66)	(9,146.89)	(10,062.19)	(1.09)
Observations	393	393	393	393
R^2	0.59	0.74	0.74	
Adjusted R^2	0.58	0.74	0.73	
Log Likelihood				$-3{,}489.29$
θ				21.36^{***} (1.52)
Akaike Inf. Crit.				6,992.59
Note:				*p<0.1; **p<0.05; ***p<0.01
	This regression output shows the results using 3 different dependent variables and two models			

[1]

Table 5: Regression analysis regarding (non-) violent and total crimes and all independent variables with OLS regressions and negative binominal regression

	Dependent variable:			
	ViolentCrimeRate	NonViolentCrimeRate	CrimeRate	
	OLS	*OLS*	*OLS*	*negative binomial*
	(1)	(2)	(3)	(4)
FoundationsDensity100k	3.89***	19.94***	23.84***	0.002***
	(0.84)	(6.33)	(6.95)	(0.001)
FlowRate	0.01**	0.08**	0.09**	0.0000***
	(0.01)	(0.04)	(0.04)	(0.0000)
TurnoutPercentage	−13.84***	−30.33	−44.17	−0.002
	(3.58)	(26.84)	(29.43)	(0.003)
BelieversRate	−0.07***	−0.71***	−0.78***	−0.0001***
	(0.01)	(0.06)	(0.07)	(0.0000)
MarriageRate	−0.29**	−1.43	−1.72*	−0.0000
	(0.12)	(0.92)	(1.01)	(0.0001)
MaleRate	−103.94***	−926.43***	−1,030.38***	−0.11***
	(20.23)	(151.70)	(166.36)	(0.02)
YouthRate	21.69***	35.43	57.11**	0.01**
	(3.38)	(25.34)	(27.79)	(0.003)
UnemployedPercentage	58.63***	816.05***	874.68***	0.10***
	(4.98)	(37.31)	(40.92)	(0.005)
Constant	12,778.81***	119,640.60***	132,419.40***	20.35***
	(1,314.43)	(9,856.26)	(10,808.72)	(1.17)
Observations	393	393	393	393
R^2	0.62	0.75	0.75	
Adjusted R^2	0.61	0.75	0.74	
Log Likelihood				−3,478.45
θ				22.56*** (1.60)
Akaike Inf. Crit.				6,974.89

Note: *p<0.1; **p<0.05; ***p<0.01
This regression output shows the results using 3 different dependent variables and two models

[1]

References

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach, 76 J. Pol. Econ. 169 (1968)." *Journal of Political Economy*.

Chan, Chung-hong, Geoffrey CH Chan, Thomas J. Leeper, and Christopher Gandrud. 2015. *Rio: A Swiss-Army Knife for Data I/O*. http://CRAN.R-project.org/package=rio.

Gandrud, Christopher. 2015. *Repmis: Miscellaneous Tools for Reproducible Research*. http://cran.r-project.org/package=repmis.

Hanslmaier, Michael. 2015. "Mehr Sozialkapital–weniger Kriminalität? Eine Makrosoziologische Längsschnittanalyse Bayerischer Landkreise." In *Empirische Forschung über Kriminalität*, 361–404. Springer.

Lederman, Daniel, Norman Loayza, and Maria Menedez. 2002. "Violent Crime: Does Social Captal Matter?" *Economic Development and Cultural Change* 50 (3): 509–39.

OECD. n.d. doi:http://www.oecd.org/insights/37966934.pdf.

Putnam, Robert D. 2001. *Bowling Alone: The Collapse and Revival of American Community*. Simon; Schuster.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2015. *Tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. http://CRAN.R-project.org/package=tidyr.

Wickham, Hadley, and Winston Chang. 2015. *Ggplot2: An Implementation of the Grammar of Graphics*. http://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley, and Romain Francois. 2015. *Dplyr: A Grammar of Data Manipulation*. http://CRAN.R-project.org/package=dplyr.