Social interventions connect to the origin of the population

Samuel Schiess[1] and Joschka Geissler[1]

[1]EPFL

January 12, 2018

Introduction

The world human population is steadily increasing. In developed countries, > 70 % of the population live in cities or urban areas (Bettencourt et al., 2007). With the increase of the population, humankind is facing new challenges and existing problems become worse. Not only problems of famine, water scarcity or increased energy demand can occur, also socially problematic situations arise when many people live in the same spots (Glaeser and Sacerdote, 1999). With the density of humans living in the same area, also psychological problems occur. The social pressure rises and the concurrence in every domain is increasing, as so many people in the near areas specialize in the similar disciplines. On the other hand, people need to choose a lifestyle, find their way of living and are responsible for their own well being. The pressure and responsibility can often end in psychological problems, which can lead to domestic violence, self-hatred and even suicide attempts. For actions that do not respect the law, the police is forced to intervene. In our digitized world everything and also those social interventions are recorded.

Additionally, not only a growth of the population and the subsequent problems may be observed today. Also the mixing of the cultures and nationalities all over the world gets more important and brings new possibilities and advantages, but also new problems. In common speaking, the foreign cultures have often another treatment of psychological diseases, health, crime and violence than the native culture in for example Western Europe. This may be right in some cases, but the connection between social troubles and the percentage of foreigners in specific areas has never been clearly stated (Entorf and Spengler, 2000b).

In the case-study of the municipality of Vernier, we have access to data from the interventions due to social reasons and the demographic properties of the area under study. The context of those two variables is often taken as a political argument from right-wing parties, but a more narrow investigation will be provided in this paper. Our study area is formed by the city of Vernier with its 768 hectares and 35 300 inhabitants, situated in the Canton of Geneva and thus a part of the metropolitan area of the city of Geneva (Vernier, 2017). The main focus will be set to investigate the spatial correlation between those two variables. In addition to that, other statistical measurements will be provided, based on the data set and the area given.

Hypothesis

In the first step we want to investigate the existence and the properties of hotspots in the region of Vernier regarding the two factors social interventions by the police and the origin of the population. Secondly, the relation between those two variables is of interest and their spatial correlation will be investigated. The expected facts are, that both parameters underlie a spatial distribution and hotspots exist and that the positions of the hotspots and thus the parameters are positively correlated to each other.

Data

We use population data from STATPOP of the year 2015 which is annually surveyed by the Federal Statistical Office of Switzerland for the households of the country (Geneva, 2017). From this data, quantitative information about the total population, the number of Swiss, non-Swiss and non-European inhabitants have been used. The second data set used is the data published by the cantonal police department of Geneva. The data set provides all police interventions due to social reasons in the area under study from 2014 to 2017 (SITG, 2017).

Methods

All spatial and statistical analyses have been performed with the software Geoda (Geoplan, 2017). QGIS (Version 2.18.) (QGIS, 2018) has been used for the visualization.

Having the two data sets described, we were able to perform a first statistical analysis. Therefore we performed the processing chain shown in Figure 1. The first step was to create a grid over the study area of vernier with a cell-size of 100 m x 100 m. For each cell, the total number of interventions has been summed up and divided by the total number of the population in this specific cell. Additionally, the number of Swiss, Non-Swiss, and Non-European inhabitants has been computed for each cell. To get the percentage of the different origins, those numbers have also been divided through the total number of the population within the cell. With the obtained variables, we were able to perform a first Linear Regression (Non-Swiss/Interventions, Non-European/Interventions). In addition to the Linear Regression, we also computed the Raw Rate. Therefore, the number of interventions divided by the total population was divided through the percentage of Non-Swiss and Non-European inhabitants within each cell.

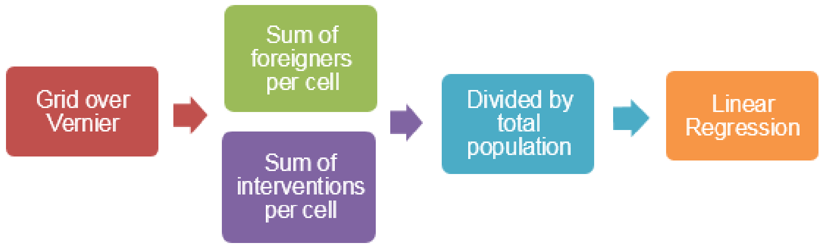


Figure 1: Processing chain for Raw Rate and Linear Regression

In order to find the socalled hotspots of our variables we had to perform a spatial analysis. The workflow for this spatial analysis is shown in Figure 2.

Figure 2: Processing chain for Spatial Rate Smoothening

First, a Queen's 2 weight file was calculated for the grid of vernier. Figure 3a shows the idea of this weight file. Each cell with a distance closer or equal to two is taken into account (green) for the computation of the value of the cell of interest (red). The values of the white cells do not have an influence on the calculation of the red value. For instance, if we want to calculate the spatial rate smoothed value for the red cell, we would divide the sum of the first parameter (e.g. Number of Interventions) of all green colored cells through the sum of the second parameter (Total Population) of all green cells. Several "Spatial Rate Smoothed"-rates (SRS) were computed (Percentage of Non-Swiss inhabitants per population, Percentage of Non-European inhabitants per population, Number of Interventions per population). All results have been plotted, visually analysed, discussed and compared to the raw rate which has not been spatially smoothed.

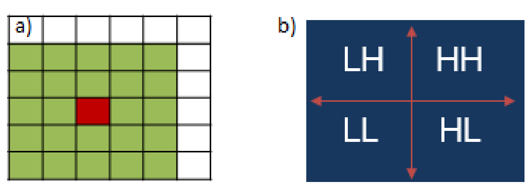


Figure 3: Spatial Analysis; a) Concept of the Queen's 2 weight file; b) Concept of the Local Moran's I

To investigate dependencies between the variables the Local Moran's I-analysis was performed in addition to the raw and the spatially smoothed rates. The Local Moran's I provides a possibility to check spatially distributed attributes in its context. As we are interested in the connection between the interventions and the origin of the population, we have chosen the Bivariate method. It enables a comparison of the values for the variables with its neighbors and presents the correlation of the two variables at the same time. When working with *Local* Moran's I, the value of each cell is associated to its specific location and not with the global context.

Local Indicators of Spatial Association (LISA) measures the association for each spatial unit and identifies the type of spatial correlation. For instance, the Bivariate Local Moran's I gives an indication of the sign of linear association (positive or negative) between the averaged value of the first variable at a given location and the averaged value of another variable at neighboring locations (Ali Akbar Matkan, 2013).

As Figure 3b shows, cells with relatively high values for both variables end up in the top-right quadrant "HH". Cells with relatively low values for both variables end up in the "LL"-quadrant (positive association) . Cells with a high and a low value for the variables are located in the "LH" or "HL"-quadrant respectively (negative association). In case the variables of a cell are not significantly high or low the cell is categorized as non significant using a threshold for significance (p-value).

Results

Regarding the hotspots, a simple representation of the variables over the area has been applied. To observe existing neighboring effects and to include a smoothening filter, the SRS was applied for both parameters using a Queen's 2 weights file. To be able to compare the results for raw and smoothed data directly, the images are represented aside each other. The cells are organized in 5 quantiles, with the darkest being the highest and the lightest being the lowest quantile. Figure 4 shows the interventions per population, while Figure 5 represents the percentage of foreigners per cell.

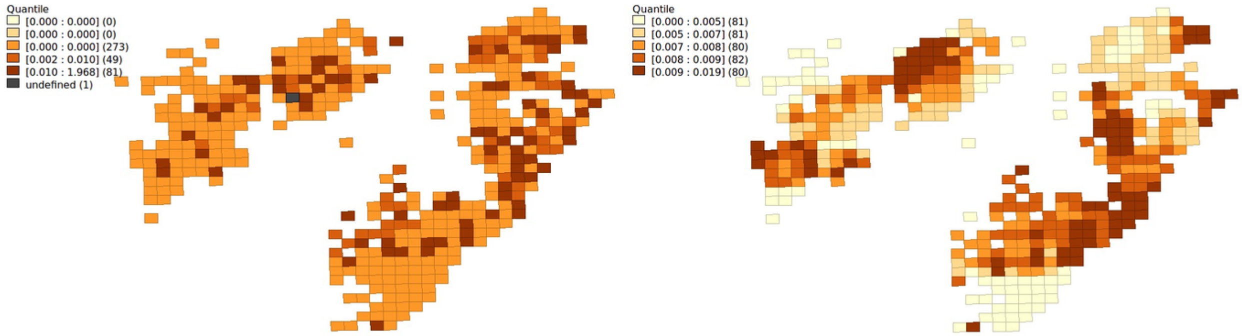


Figure 4: Shows the cells of the municipality of Vernier for the variable interventions divided by the population of each cell. On the left the raw data, without treatment is represented, while the right map shows the smoothed data according to SRS. For the raw data, the legend shows the number of interventions per person per cell, while the smoothed data on the right has another unit and has to be looked at relatively to its values and not be directly compared to the raw data.

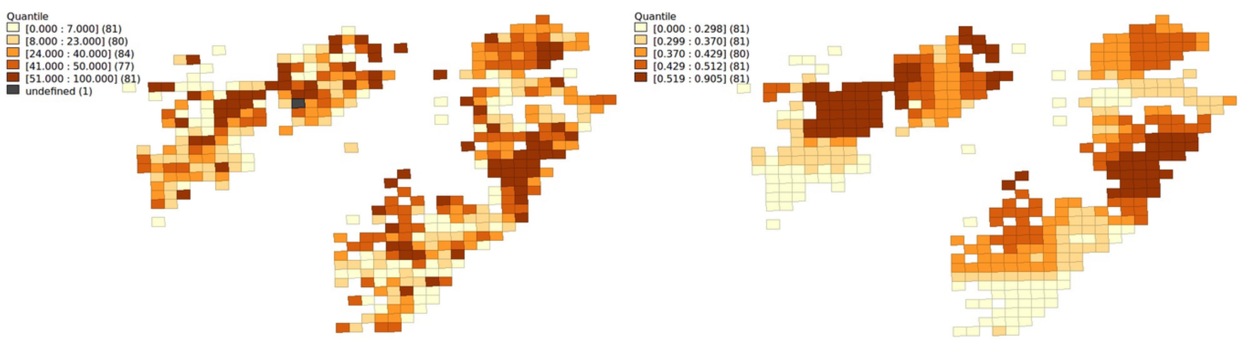
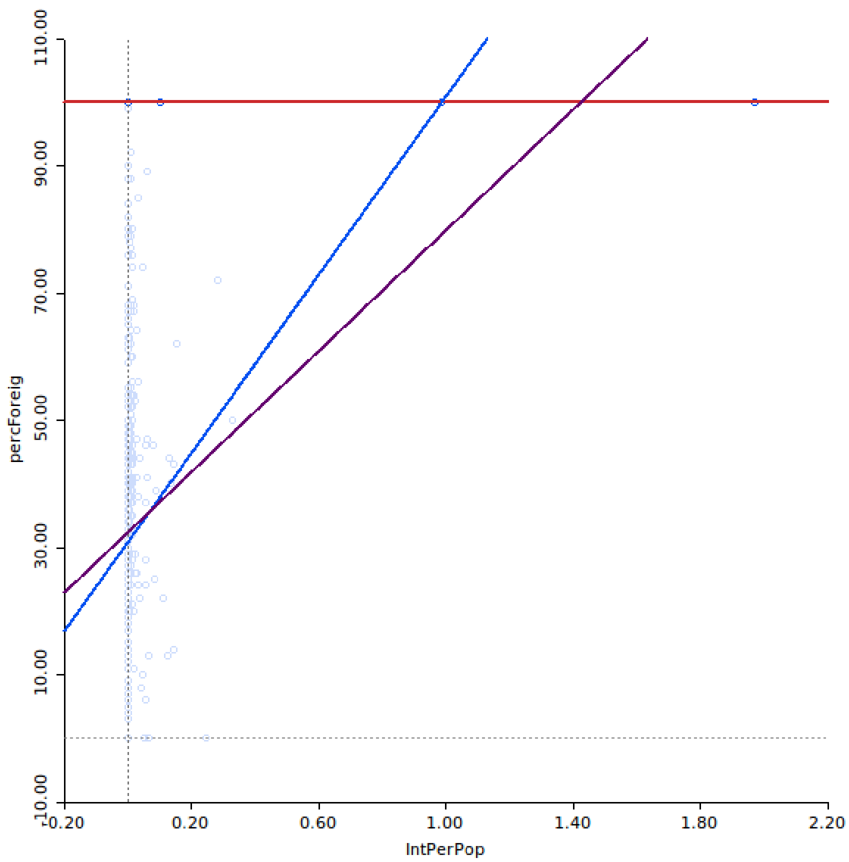


Figure 5: Shows the cells of the municipality of Vernier for the percentage of foreigners living in a cell. T left map shows the raw data, while the right map represents the smoothed data with the two hotspots in the northern part. For the raw data, the legend shows % of the total population, while the smoothed data on the right has another unit and has to be looked seperately.

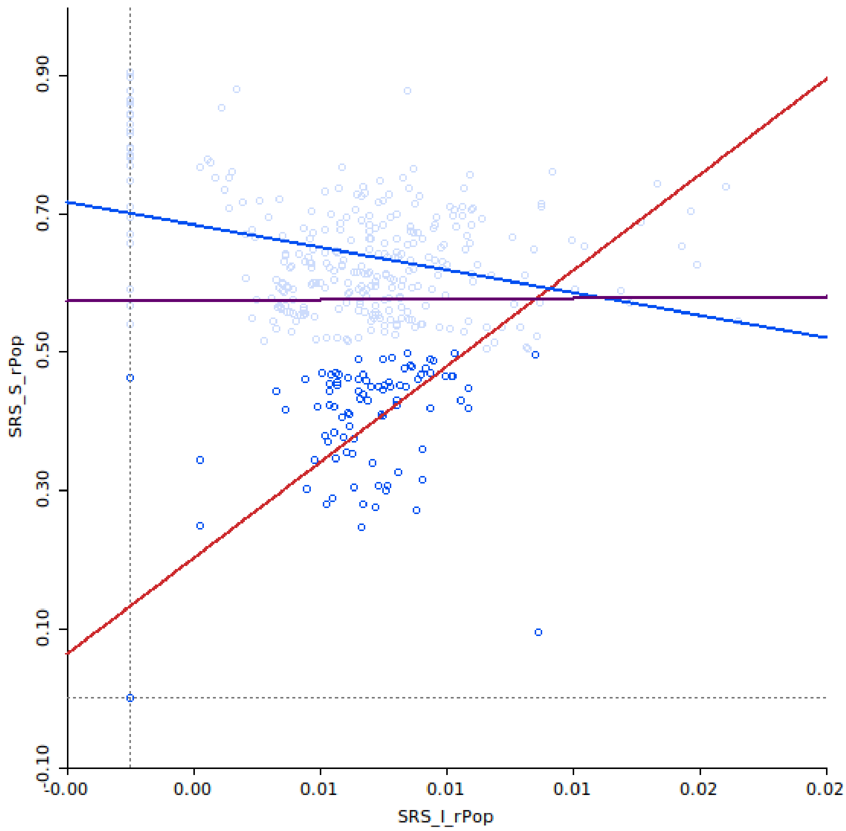
To investigate the relation between the two parameters, we first used common statistical methods such as a regression of the data. In the following, the two scatterplots for the two parameters are presented. The first shows the regression of the percentage of foreigners and interventions and its statistics for the raw data, while the second shows the variable interventions with the percentage of Swiss inhabitants for the smoothed data according to the SRS and the Queen's 2 weight file. in Figure 6 the points with a 100% foreign population are selected and the blue line shows the regression without those ten points. In the second scatterplot in Figure 7, the cells with $< 50\%$ Swiss citizens are selected and the blue line shows the regression when ignoring those.



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
403	0.053	32.439	1.252	25.915	0.000	47.590	10.033	4.743	0.000
10	1.000	100.000	0.000	100.000	0.000	0.000	0.000	100.000	0.000
393	0.009	30.981	1.233	25.122	0.000	70.124	36.382	1.927	0.055

Chow test for sel/unsel regression subsets: distrib=F(2,400), ratio=31.6128, p-val=0.0000

Figure 6: Represents the regression line between the interventions due to social reasons per population and the percentage of foreigners in the cell for the raw data. The violet line and data is the regression for the whole data set, while the blue comprises the same regression when ignoring all the points with 100 % foreigners.



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
404	0.000	0.574	0.017	33.019	0.000	0.305	2.204	0.138	0.890
99	0.504	0.133	0.026	5.146	0.000	34.688	3.493	9.931	0.000
305	0.123	0.701	0.010	69.866	0.000	-8.128	1.249	-6.509	0.000

Chow test for sel/unsel regression subsets: distrib=F(2,400), ratio=460.9262, p-val=0.0000

Figure 7: Shows the regression line between the interventions due to social reasons per population and the percentage of Swiss inhabitants in the cell for the smoothed data using SRS and the Queen's 2 weight file. The violet line is the regression of the whole data and the blue one is the regression when ignoring all points with a Swiss population of < 50 %.

To investigate the spatial correlation between the two variables, we applied the Local Moran's I, using the same Queen's 2 weight file as before. The results are plotted on an Open Street Map using the software QGIS. 28% of the cells have a significant relation, which is supporting our hypothesis. 42% of the cells do not have a significant correlation between the two variables and are therefore marked in lightyellow.

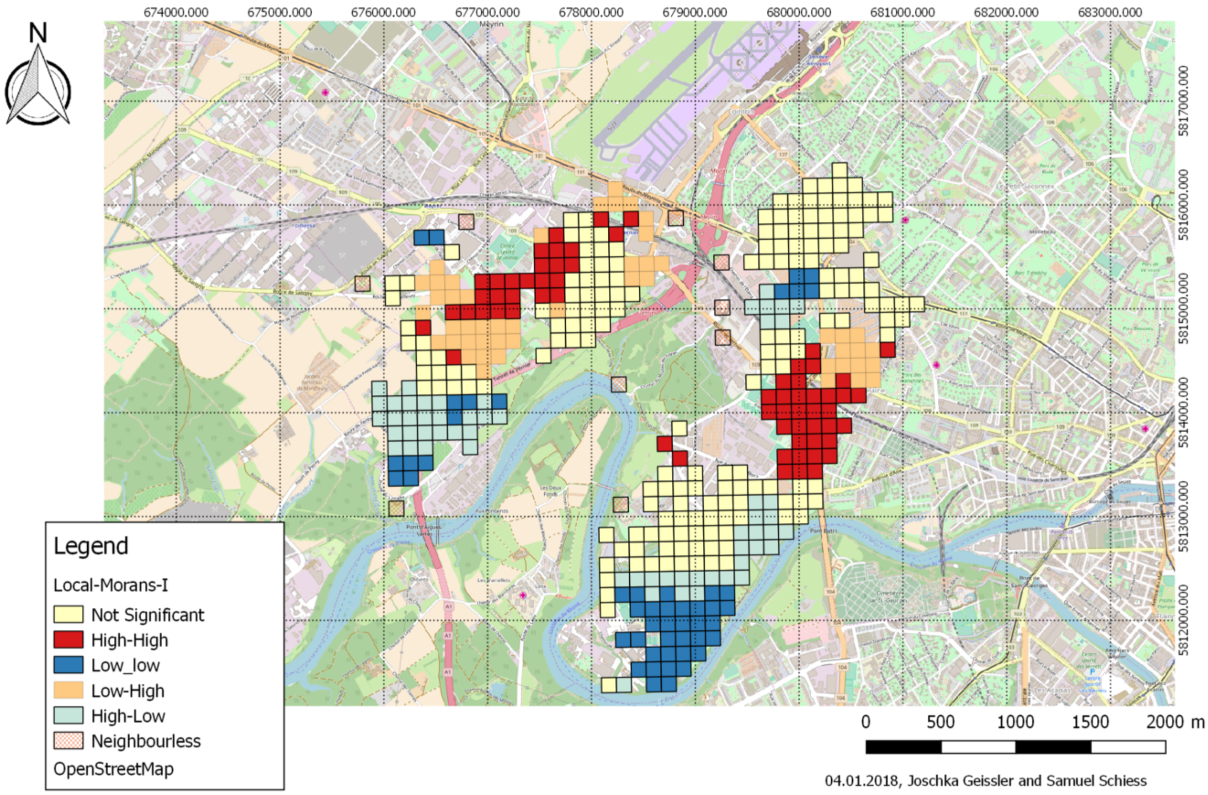


Figure 8: The map shows the cluster representation of the Local Moran's I using the Queen's 2 weight file. A significance of 0.001 was assumed. Two hotspots of High-High relation can be spotted and one large area of Low-Low relation is situated in the south of the municipality. The background map has its source in the Open Street Map integrated in QGIS.

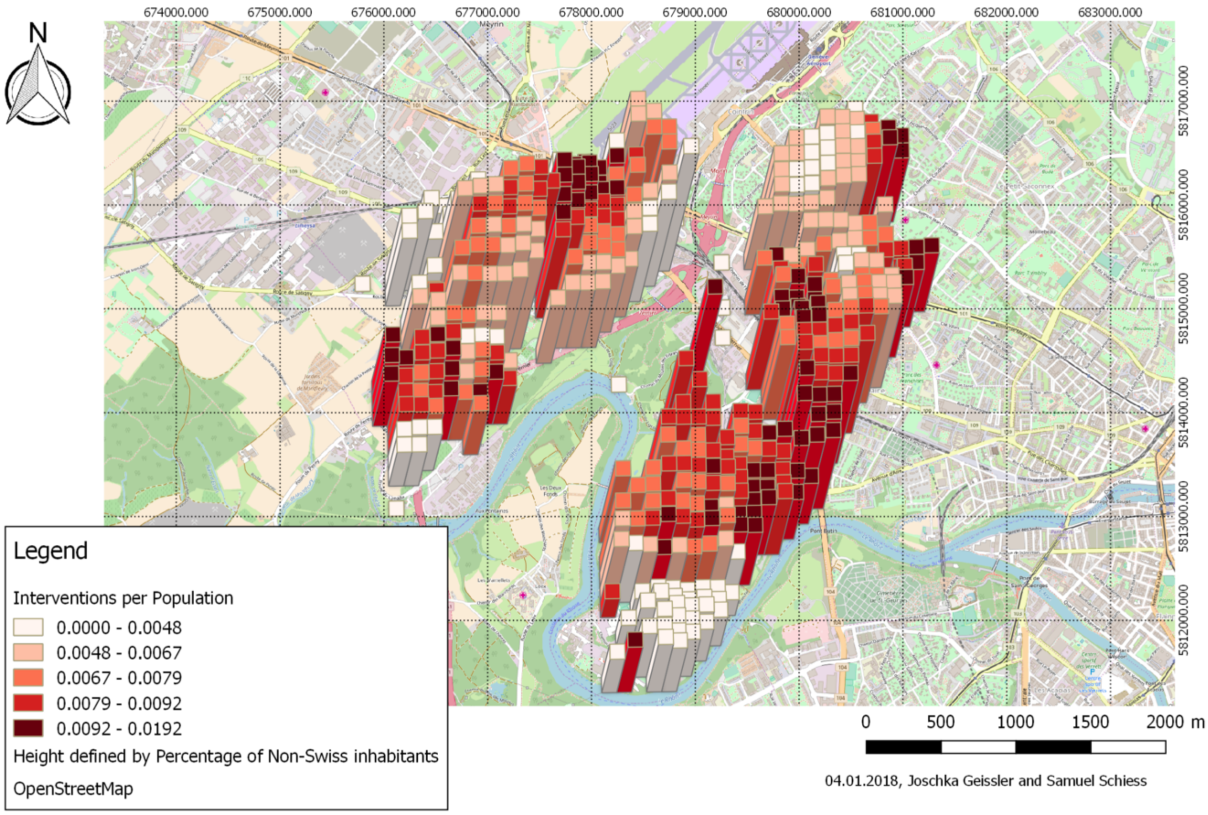


Figure 9: The map shows a spatial analysis comparing the percentage of non-Swiss inhabitants with the number of interventions per cell. Both variables have been smoothed with Queen's 2 contiguity. Compare the map with the right map of Figure 5. The background map has its source in the Open Street Map integrated in QGIS.

Discussion

Both hypotheses were proved considering the smoothed data of the SRS analysis. Nevertheless, it is important to point out, that the smoothing effect of the SRS is quite high and a loss in precision of the data can not be excluded. Therefore, the new information varies from the provided data and the statistical check is less high than with the raw data. The strong smoothing effect has its source in the Queen's 2 contiguity. We have chosen the creation of the weight file like that, so a clear contrast between the raw data and the smoothed one is visible. Like this, we were able to point out the effect of spatial statistics and analysis compared to usual statistics. With a Queen's file of a lower order, a compromise between the smoothing and loss of data could be achieved.

For the analysis of the hotspots, the representation of the raw data and the smoothed data are already a solid base. The percentages of the foreigners living in the cells are already spatially correlated, before smoothing with SRS. Two hotspots with a share of more than 50% foreigners per cell can be detected, both with the raw and the smoothed data, as seen in Figure 5. After applying the SRS, the two hotspots are clearly separated from the other cells. This clustering of the nationalities in urban regions has already been investigated in other reports and this fact also adapts to Swiss cities (Adelman et al., 2016).

For the interventions divided by the population, some hotspots are established, although not as clear as for the percentage of foreigners as to be seen in Figure 4. On the other hand, the raw data would not show clusters in any quantile, which warns us, that the smoothed data should be handled with a critical mind. The literature also states a clustering of social interventions, although the exact distinction between common crime due to social problems is not clear (Massey et al., 1991). To further proof this fact, our study area is certainly not large enough. Another aspect is the restricted period of time that the data of the interventions is representing. A time series over a longer time period could improve the result.

The boxplots in Figure 6 and 7 do not clearly show a correlation between the variables, neither for the raw data nor for the smoothed. Also when playing with the selection of the cells, no clear correlation can be stated. When normalizing the data, the slopes do not show the expected slope of 1, but much smaller values. In addition, the R^2 do never exceed 0.1, which makes the regression not reliable, with a huge error that can not be explained by the regression curve. Thus, with a statistical analysis of the connection between interventions and percentage of foreigners no clear correlation can be detected.

For the further investigation, a spatial analysis was adapted using the Local Moran's I. When comparing the clustering of the Local Moran's I in Figure 8, clear tendencies exist in building two hotspots for the High-High correlation and also one for the Low-Low connection. Also in this case, the smoothing effect of the SRS method must be taken into account, which helps to build clusters from the variables. With only 28% of the area, resulting in 113 cells showing a significant relation between the two variables, the effect of the variables on each other are rather weak. The clusters of the High-High connection lie in those areas where it was to be expected according to the hotspots of the smoothed data in Figure 5. The spatial distribution of interventions and foreigner population would be expected as more correlated, according to general speaking (Vimentis, 2017). Although scientifically seen, this connection is not clearly proofed (Entorf and Spengler, 2000a) and Adelman et al. even states the opposite(Adelman et al., 2016). The reason for the thrown hypothesis in this case can also be the same as for the hotspots, as the area is not large enough and the data not that comprising. Also, the interventions due to social reasons do not necessarily correlate with crimes in general.

The two important living areas in Vernier, Lignon and Avanchets, do not specially appear in neither of the analysis. The reason could be, that the interventions are divided by the population living in the cell. But also in the analysis of the percentage of foreigners per cell, the areas are not showing special properties regarding the variable. The region of Lignon has become a popular living area and therefore, the prices do not specially attract foreigners or other social classes.

When looking at the results overall, it can clearly be said, that the correlation between the percentage of foreigners living in an area and the social interventions by the police is rather weak. Other social properties, such as education, wealth, stress etc. must have a more important influence in this region and could be investigated in a further study with more input data regarding those topics.

References

Robert Adelman, Lesley Williams Reid, Gail Markle, Saskia Weiss, and Charles Jaret. Urban crime rates and the changing face of immigration: Evidence across four decades. *Journal of Ethnicity in Criminal Justice*, 15(1):52–77, nov 2016. doi: 10.1080/15377938.2016.1261057. URL https://doi.org/10.1080%2F15377938.2016.1261057.

Mojgan Mirzaie Ali Akbar Matkan, Matin Shahri. Bivariate Morans I and LISA to explore the crash risky locations in urban areas. *N-Aerus 14*, 2013.

LM Bettencourt, J Lobo, D Helbing, C Kühnert, and GB West. Growth, innovation, scaling, and the pace of life in cities. *Proc Natl Acad Sci U S A*, 104:7301–6, Apr 2007.

Horst Entorf and Hannes Spengler. Socioeconomic and demographic factors of crime in Germany. *International Review of Law and Economics*, 20(1):75–106, mar 2000a. doi: 10.1016/s0144-8188(00)00022-3. URL https://doi.org/10.1016%2Fs0144-8188%2800%2900022-3.

Horst Entorf and Hannes Spengler. Socioeconomic and demographic factors of crime in Germany. *International Review of Law and Economics*, 20(1):75–106, mar 2000b. doi: 10.1016/s0144-8188(00)00022-3. URL https://doi.org/10.1016%2Fs0144-8188%2800%2900022-3.

Canton Geneva. Canton Geneva 2017. https://www.ge.ch/statistique/domaines/01/01$_0 2_2 /methodologie.asp$, 2017.U Accessed on Tue, December 12, 2017.

Geoplan. Geoda Project. https://geoplan.asu.edu/geodacenter-redirect, 2017. URL https://geoplan.asu.edu/geodacenter-redirect. Accessed on Thu, January 11, 2018.

Edward L. Glaeser and Bruce Sacerdote. Why is There More Crime in Cities? *Journal of Political Economy*, 107(S6):S225–S258, dec 1999. 10.1086/250109. URL https://doi.org/10.1086%2F250109.

Douglas S Massey, Andrew B Gross, and Mitchell L Eggers. Segregation the concentration of poverty, and the life chances of individuals. *Social Science Research*, 20(4):397–420, dec 1991. 10.1016/0049-089x(91)90020-4. URL https://doi.org/10.1016%2F0049-089x%2891%2990020-4.

QGIS. QGIS Project. https://www.qgis.org/en/site/, 2018. URL https://www.qgis.org/en/site/. Accessed on Thu, January 11, 2018.

SITG. Catalogue — SITG. http://ge.ch/sitg/sitg$_c atalog/sitg_d onnees?keyword = \&distribution = tous\&datatype = tous\&topic = society\&service = tous$, 2017. URL. Accessed on Thu, January 11, 2018.

Vernier. Vernier 2017. http://www.vernier.ch/fr/portrait/presentation/, 2017. URL http://www.vernier.ch/fr/portrait/presentation/. Accessed on Tue, December 12, 2017.

Vimentis. Criminality 2017. https://www.vimentis.ch/content/docs/PA$_A uslaenderkriminalitaet_E ndversion.pdf$ Accessed on Tue, December 12, 2017.