Spatial Research Report

Jonas Lauri

December 9, 2022

Abstract

How are the rates of urbanization affected by the presence of higher institutions? This is a very relevant question as higher rates of urbanization have obvious effects on society, which will be briefly introduced. Using simple relations, this project examines any link between these two. Specifically, we will look at Sweden and try to correlate the population and number of higher institution entrants per county. The resulting regressions show that the potential correlation is quite weak. The reasons for this result are problems with the simple methodology and the data. If one wants to continue this research in the future, in would be good to get finer data or augmentations to the data set as only better methodology can only lead so far.

Introduction

The world is quickly become more and more urbanized by the day. In Sweden most the population is already living in cities. Of course, what is classified as an urban area depends on what definition one has of a city. For example, if one classifies a city as being constituted by at least 200 people and live close to each other, then Sweden in 2018 was 87 percent urbanized. (There are some other rules as well but the 200 persons one is the most important in this context). If one where to increase the threshold to 10 000 persons then number would be 63 percent and for 100 000 person cities then it would be 32 percent [1]. With these numbers in mind, we see that while most people live in urban areas, there is still a lot of big city urbanization possible.

Lets say there is a lot of urbanization coming, in that case, what are the effects of this increased urbanization? This is of course a very complex question as the effects are different at different locations but in Akpans article, *The impact of urbanization and institutions of higher education on Houston Texas' Third Ward Community*, he found that the establishment of higher institutions did actually negatively impact the neighbouring communities [2]. In the conclusion he speculated that to get positive outcomes, just establishing higher institutions is not enough, one also needs to make sure that for example the political will [to make sure there are positive outcomes] is there as well. These problems with urban areas, especially of low socioeconomic class, are concurred in Nukhet, N. Gamze and Yuksels article, *The importance of urbanization in education*, which specifically looked at the outcomes in education when comparing rural and urban areas [3]. However, while the same problems occurring in cities were discussed, they also found that urban areas had higher performance in education.

So what conclusion can one draw from these studies? The answer is that urbanization and education are both part a such a big system which makes it is hard to draw any definite conclusions. However, any potential links between urbanization and educational institutions are interesting as the world will keep urbanizing.

Data and methods

This report will focus on the effects of existing higher institutions on the rate of people moving in. To do this we need population data, both national and per county, and also data on where the higher institutions are situated.

Population data

The population data is taken from SCB, *Statistiska centralbyrån* or *Statistics Sweden*, and we will be looking at *Inflyttningar* and *Invandringar*, which are move-ins and immigration to get amount of people moving into each county each year [4][5]. The total population of each county is also from SCB, and has both the total for each year and the net for each year [6].

Institution data

As for the data for the higher institutions, we will use data from *Universitetskanslersämbetet*, UKÄ, or in english, *The Swedish Higher Education Authority* [7]. We will look at the county which the institution is located and the number of new entrants. In figure 1 we can see the number of unique higher institutions per county.

While not all educations are equal, (some programmes have higher levels of dropouts) in this report we will just take a look at how many entrants there were in 2018 and assume that the number of entrants scales linearly with total population size. This is of course not the case as the popularity of different universities change over time. Additionally, Sweden recently got large influx of immigrants which skewes the data. However, these assumptions will make the analysis possible and possible effects will be further discussed later. In fact, we will not assume anything about the differences between different universities, university colleges or art, design and music academies, which will analyzed. Futhermore, some big universities have campuses in multiple cities, which will not be considered.

Method

First we need to establish the model which we will analyze. As we would expect the relationship to be quite simple we should implement a simple model. However, how to implement this is a very long winded process. First we need a way to guess how the growth of each county would look if there was no need to move for education proposes. This would correspond to the linear regression which can be seen in figure 2.

Then we sum the populations over each year for each county. From this we get the population of each county by scaling with the population proportion. Then get the enrolled student proportion by looking at specifically 2018, as the data is from 2018. Finally we do the regression according the following equation.

$$\beta_i = [data_i - pop_i(year)] \cdot enrolled\_student\_population_i \cdot year \qquad (1)$$

where *data* is the real population from SCB, *pop* is the estimated population based on Swedens total growth rate as mentioned above and i are the different counties. The reason why we structure the regression in this way is that if data is larger than pop than we would expect an effect caused by the presence

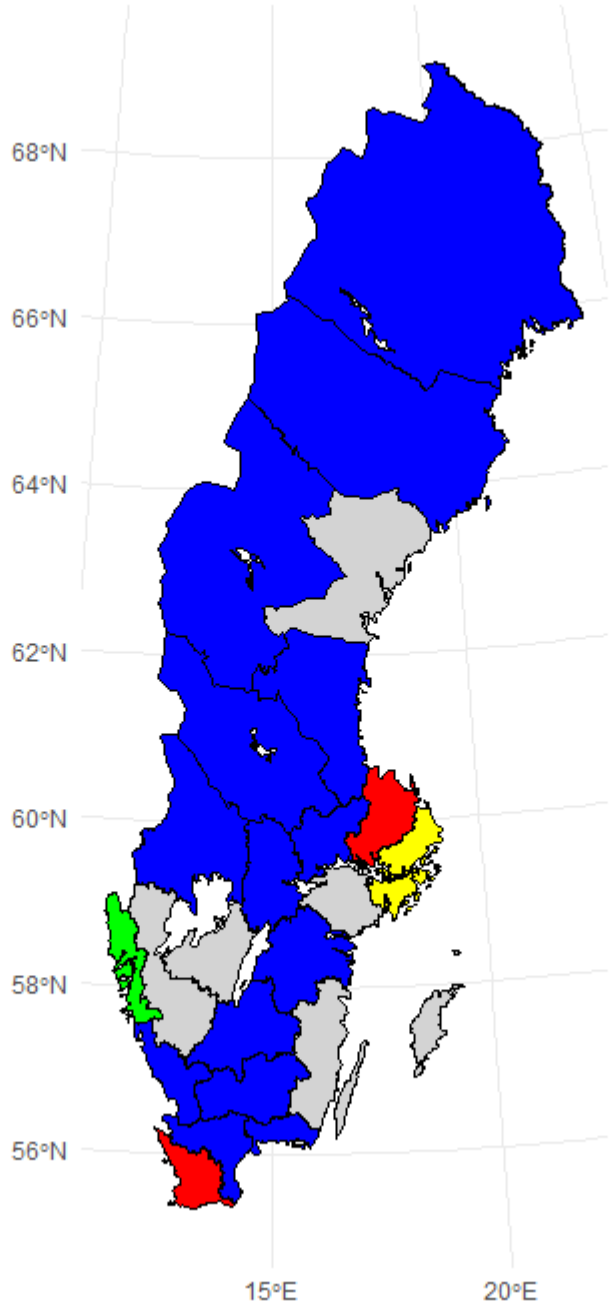

Figure 1: A map of Sweden where each county has been colored corresponding to the number of institutions within its borders. Grey, blue, red, green and yellow correspond to zero, one, two, four and thirteen institutions.

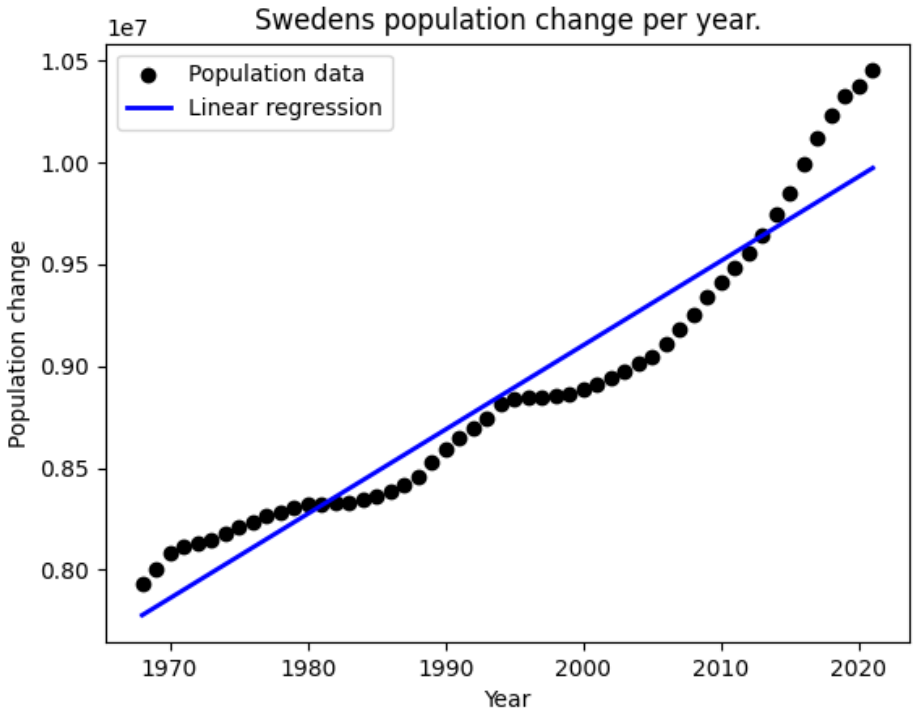

Figure 2: The resulting linear regression from applying a linear model to Swedens population. This resulting regression has a $R^2 = 0.92$.

of higher institutions. Notice that equation 1 has no constant factor. The linear regression is made with the scikit-learn package [8].

Hypothesis

We want to compare the population growth of cities with higher institutions as compared with the country as a whole to see if the presence of these institutions affect the growth rates. The hypothesis is that we have a positive correlation between the presence of higher institutions and growth rates.

Results

The results of the regression can be seen in table 1. The counties which does not have any higher institutions are removed from this table.

Discussion

As stated in table 1, the results show that there is the correlation is quite small. This means that there is basically no information to gain from the occurrence of higher institutions when analyzing. One big problem is that the population data used in the final regression is very sporadic. This can be seen by comparing the total population to the population change per year, as seen in figure 3.

Another part that makes the result unreliable is all the assumptions made in this project. For example, the growth rate of each county by just scaling with the population proportion is not very realistic. The data only going back to 1968 is also a problem. Then, any university which was established before 1968 becomes absorbed into the data. A hard problem to solve.

The small positive correlation we did get is also hard to comment on. It is only a slight correlation and that could be because big cities tend to get bigger.

Continued research should focus on making less assumptions and making use of more spatial relations, which is not feasible in a small project such as this.

County	*enr_stud_prop*	β	R^2
01 Stockholms län	0.00811391	19.75241484	0.20217139
03 Uppsala län	0.02303682	10.86932937	0.23751877
05 Östergötlands län	0.01102727	-1.25854872	0.12012791
06 Jönköpings län	0.00778771	-1.24594884	0.19763747
07 Kronobergs län	0.02566463	-1.8736142	0.1456583
10 Blekinge län	0.0060745	-1.03168297	0.24172588
12 Skåne län	0.0095363	5.58858056	0.13990624
13 Hallands län	0.00640652	2.08755749	0.24741341
14 Västra Götalands län	0.00746865	-0.26580286	0.00123998
17 Värmlands län	0.00888156	-3.38804283	0.25133179
18 Örebro län	0.0083705	-1.86207569	0.20637931
19 Västmanlands län	0.00857887	-2.39161047	0.22953125
20 Dalarnas län	0.00727739	-2.49245103	0.24335707
21 Gävleborgs län	0.00624679	-2.65316521	0.2499586
23 Jämtlands län	0.01558182	-3.02157586	0.24607502
24 Västerbottens län	0.01610193	-2.01612307	0.21718828
25 Norrbottens län	0.00922167	-3.58422575	0.24770809

Table 1: The results from the linear regressions. We see that while there is information to gain by using the enrolled student data, it hardly matters as the R^2 are very small for all regressions. The R^2 is positive though, so there is a small merit to using this data.

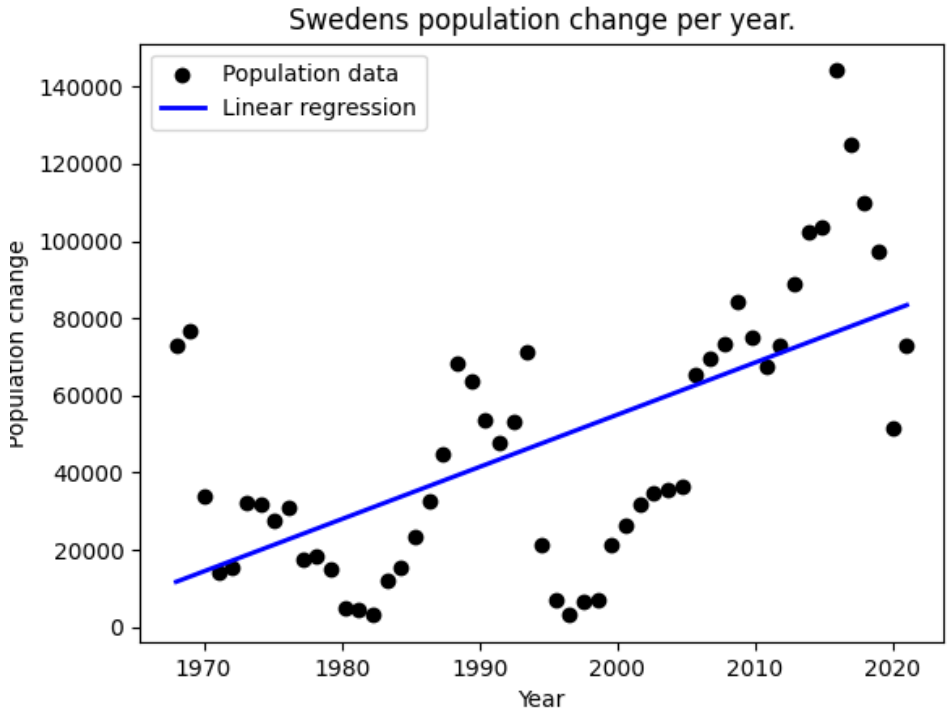

Figure 3: The resulting linear regression from applying a linear model to Swedens population change per year. This resulting regression has a $R^2 = 0.38$. (Total population regression had a $R^2 = 0.92$)

References

1. Sweden, Statistics. *Localities and urban areas 2018*. 2018. URL: https://www.scb.se/en/finding-statistics/statistics-by-subject-area/environment/land-use/localities-and-urban-areas/pong/statistical-news/localities-and-urban-areas-2018/ (visited on 12/07/2022).

2. Akpan, Akpan I. "The impact of urbanization and institutions of higher education on Houston Texas' Third Ward Community". In: *Journal of Applied Sciences and Environmental Management (ISSN: 1119-8362) Vol 10 Num 2* 10 (June 2009). DOI: 10.4314/jasem.v10i2.43654.

3. "The importance of urbanization in education". In: *The Eurasia Proceedings of Educational & Social Sciences (EPESS)* 5 (2016), pp. 232–236. ISSN: 2587-1730.

4. Sweden, Statistics. *Flyttningar efter region, ålder och kön. År 1968 - 1996*. URL: https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101J/Flyttningar/ (visited on 12/07/2022).

5. — *Flyttningar efter region, ålder och kön. År 1997 - 2021*. URL: https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101J/Flyttningar97/ (visited on 12/07/2022).

6. — *Folkmängden efter region, civilstånd, ålder och kön. År 1968 - 2021*. URL: https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101A/BefolkningNy/ (visited on 12/07/2022).

7. *Higher education institutions in Sweden 2020 status report*. 2020. URL: https://english.uka.se/about-us/publications/reports--guidelines/reports--guidelines/2020-08-26-higher-education-institutions-in-sweden---2020-status-report.html.

8. Pedregosa, F. et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

Code

The code can be found at: https://github.com/BotLauri/TOUDAI_SPATIAL.