

# Information on sampling

## SNSF Ambizione Project „The neighborhood in the cloud”

Christoph Zangger

### Starting Position

The project „The neighborhood in the cloud“ (<https://data.snf.ch/grants/grant/201694>), supported by Swiss National Science Foundation, addresses the question of how networks with a spatial context form both on- and offline and how these, together with other geographical and social settings, influence individual educational and labor market opportunities as well as political participation. In order to obtain results that are as robust and generalizable as possible, the project will use a variety of data sources. These include a survey of around 9,000 people living in Switzerland, who are interviewed at two points in time. Existing data neither permit the identification of small-scale neighborhoods required for the project nor do they possess the necessary relevance in terms of their content, making this additional survey necessary.

### How were the sampled municipalities selected?

The target number of 9,000 interviewed (or contacted) people allows conclusions that are as broad as possible and generalizable to the whole of Switzerland. Essential to this is that these individuals are randomly selected. Since small-scale social relationships between immediate neighbors are of primary importance, we cannot use a simple random sample of people or households. Instead, a total of 300 coordinates/addresses are randomly drawn and the 30 geographically most proximate households to these coordinates are invited to participate. However, such an approach is not feasible using the sampling frame of the Federal Statistical Office, and the corresponding contact data must be requested directly from the residents' offices.

In order to minimize the amount of offices to be contacted and at the same time to guarantee the representativeness of the drawn sample, a total of 86 municipalities were selected, 20 of which are municipalities with an urban character and at least 20,000 inhabitants, another 20 are urban in character with less than 20,000 inhabitants, 25 are rural municipalities and 21 municipalities have an "intermediate" character according to the classification of the Federal Statistical Office. The drawing probability within the spatial typology was proportionate to the size of the municipalities. Figure 1 illustrates the 86 randomly selected municipalities.

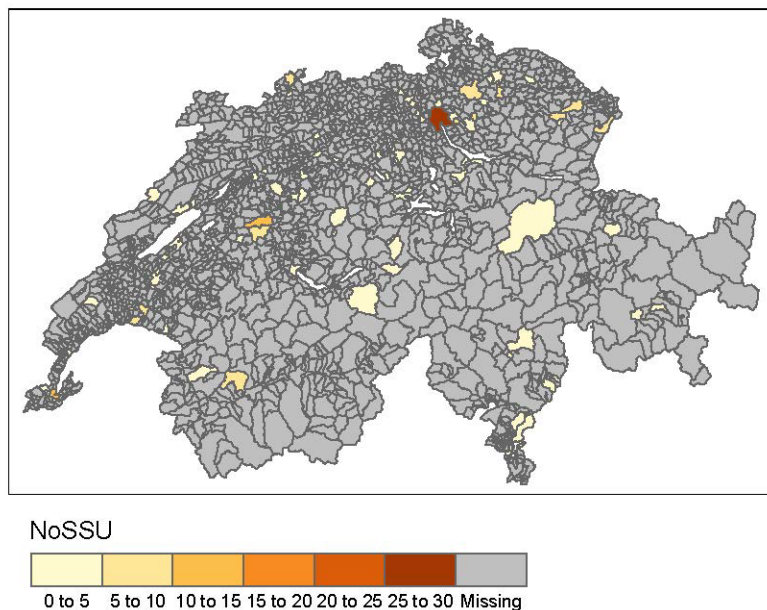


Figure 1: Randomly selected municipalities & number of coordinates to be drawn within the municipality

The 300 coordinates/addresses to be drawn within each of the 86 selected municipalities was determined in a second step. For this purpose, random selection with replacement was used, again with drawing probabilities proportional to the size of the municipality. Thereby, it was ensured that each municipality is drawn at least once. The coloring in Figure 1 indicates the number of coordinates/addresses to be drawn.

### **How were the individual respondents selected?**

Once the municipalities to be contacted and the number of people to be interviewed in each of them (number of coordinates to be drawn x 30) were determined, the next step was to randomly select the coordinates (target points of the sample) around which the next 30 households would be drawn within the municipalities. In order to take into account the population density in the relevant municipalities while drawing the target points, the official building register was consulted. Thereby, only addresses of buildings with exclusively residential use (GKAT 1020) and residential buildings with auxiliary use (GKAT 1030) are considered. Buildings with partial residential use (e.g., homes or hospitals; GKAT

1040) were excluded, as well as all addresses of non-residential buildings (GKAT 1060), special buildings (GKAT 1080), and temporary housing (e.g., construction containers; GKAT 1010).

Simple random sampling from this database will result in addresses in densely populated areas (many buildings on the same street) being drawn more frequently than, for example, a remote alpine hut. For each of the municipalities drawn in the first step, the corresponding number of target points was subsequently drawn. Surrounding these, further 29 addresses were determined according to their Euclidean distance from the drawn target address (automated procedure by using the coordinates of the respective buildings).

These clusters of 30 addresses each were then submitted to the residents' offices of the respective municipalities with the request to provide the contact details of the 30 closest households. In many cases only a few of the 30 target addresses in each group were required, since it is possible, for example, that the target drawn is an apartment building with several households. If the number of households at a particular address exceeded 30 units, they were randomly selected. The corresponding contact details were then provided to us by the residents' offices for establishing contact.

## **How is the contact data used and how is data protection guaranteed?**

The contact data as well as the collected survey data are stored on a physical server drive at the University of Bern. Only people with an employment at the University of Bern and who are directly involved in the data collection (5 individuals) can access this drive. The contact data is further saved on this server drive in a virtual encrypted drive, so even if the server is physically stolen or lost, the data is still protected. Finally, the contact data is stored only for the time of the survey (2 months) and will then be destroyed without the possibility of recovery.

For further questions please contact [christoph.zangger@unibe.ch](mailto:christoph.zangger@unibe.ch)