

# Informationen zur Stichprobenziehung

SNF Ambizione Projekt „The neighborhood in the cloud“

Christoph Zangger

## Ausgangslage

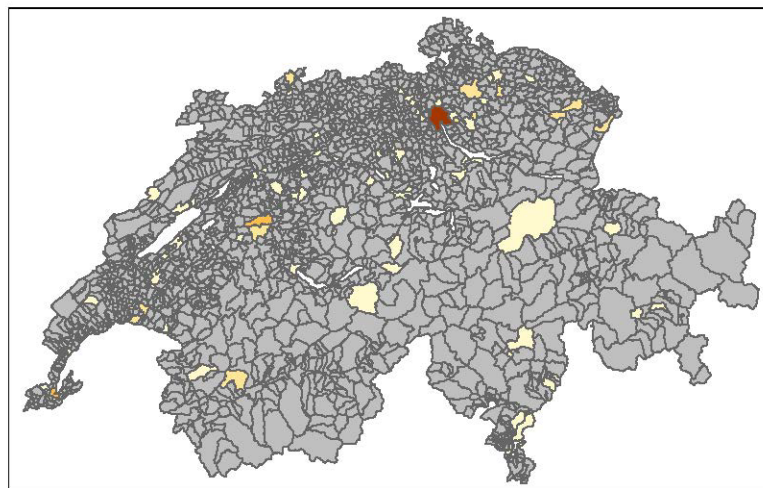
Das durch den Schweizerischen Nationalfonds geförderte Projekt „The neighborhood in the cloud“ (<https://data.snf.ch/grants/grant/201694>) beschäftigt sich mit der Frage, wie sich Netzwerke mit einem räumlichen Bezug sowohl on- wie auch offline ausbilden und wie diese zusammen mit anderen geographischen und sozialen Kontexten individuelle Bildungs- und Arbeitsmarktchancen sowie die politische Partizipation beeinflussen. Um möglichst robuste und verallgemeinerbare Ergebnisse zu erzielen, werden im Projekt unterschiedliche Datenquellen herangezogen. Hierzu zählt auch eine Umfrage unter rund 9'000 in der Schweiz wohnhaften Personen, die zu zwei Zeitpunkten befragt werden sollen. Existierende Daten erlauben dabei weder eine für das Projekt notwendige Identifikation kleinräumiger Nachbarschaften noch weisen sie den entsprechenden inhaltlichen Fokus auf, wodurch diese zusätzliche Erhebung nötig ist.

## Wie wurden die gezogenen Gemeinden ausgewählt?

Die angestrebte Zahl von 9'000 befragten (bzw. kontaktierten) Personen soll möglichst breite und für die gesamte Schweiz verallgemeinerbare Aussagen erlauben. Wesentlich hierfür ist, dass diese Personen zufällig ausgewählt werden. Da kleinräumige soziale Beziehungen zwischen direkten Nachbar:innen von zentralem Interesse sind, kommt eine einfache Zufallsauswahl von Personen nicht in Frage. Stattdessen werden insgesamt 300 Koordinaten/Adressen zufällig gezogen und die 30 geografisch am nächsten liegenden Haushalte werden zur Teilnahme eingeladen. Ein solches

Vorgehen ist indes mittels des Stichprobenrahmens des Bundesamtes für Statistik nicht durchführbar und die entsprechenden Kontaktdaten müssen von den Einwohnerämtern direkt angefragt werden.

Um die Anzahl zu kontaktierender Ämter zu reduzieren und gleichzeitig die Repräsentanz der gezogenen Stichprobe zu gewährleisten, wurden insgesamt 86 Gemeinden ausgewählt, 20 davon sind Orte mit städtischem Charakter und haben mindestens 20'000 Einwohner:innen, weitere 20 sind städtischen Charakters mit weniger als 20'000 Einwohner:innen, 25 sind ländliche Gemeinden und 21 Gemeinden weisen einen nach der Klassifikation des Bundesamtes für Statistik „intermediären“ Charakter auf. Die Ziehwahrscheinlichkeit innerhalb der Raumtypologie erfolgte dabei proportional zur Grösse der Gemeinden. Abbildung 1 zeigt die 86 zufällig gewählten Gemeinden.



NoSSU



Abbildung 1: Zufällig ausgewählte Gemeinden & Anzahl zu ziehende Koordinaten innerhalb der Gemeinde

Innerhalb der 86 Gemeinden wurden in einem zweiten Schritt die Anzahl der zu ziehenden 300 Koordinaten/Adressen festgelegt. Hierzu wurde eine Zufallsauswahl mit Zurücklegen verwendet, wiederum mit Ziehwahrscheinlichkeiten proportional zur Grösse des Ortes. Dabei wurde sichergestellt, dass jede Gemeinde mindestens einmal gezogen wird. Die Färbung in Abbildung 1 gibt dabei die Anzahl der zu ziehenden Koordinaten/Adressen an.

### **Wie wurden die einzelnen Befragten ausgewählt?**

Nachdem die zu kontaktierenden Gemeinde und die Anzahl der darin zu befragenden Personen (Anzahl zu ziehender Koordinaten x 30) festgelegt waren, wurden in einem nächsten Schritt die Koordinaten (Zielpunkte der Stichprobe) um welche die nächsten 30 Haushalte gezogen werden,

zufällig innerhalb der Gemeinden ausgewählt. Um die Bevölkerungsdichte in den entsprechenden Gemeinden bei der Ziehung der Zielpunkte mitzubedenken, wurde hierzu das amtliche Gebäuderegister herangezogen. Dabei werden lediglich Adressen von Gebäuden mit ausschliesslicher Wohnnutzung (GKAT 1020) und jene mit Nebennutzung (GKAT 1030) berücksichtigt. Gebäude mit teilweiser Wohnnutzung (z.B. Heime oder Spitäler; GKAT 1040) wurden indes ausgeschlossen wie auch alle Adressen von Nicht-Wohngebäuden (GKAT 1060), Sonderbauten (GKAT 1080) und provisorische Unterkünfte (z.B. Baucontainer; GKAT 1010).

Eine einfache Zufallsstichprobe aus dieser Datenbank führt dazu, dass Adressen in dichtbesiedelten Gebieten (viele Gebäude an derselben Strasse) eher gezogen werden als beispielsweise eine abgelegene Alphütte. Für jede der im ersten Schritt gezogenen Gemeinden wurde sodann die entsprechende Anzahl an Zielpunkten gezogen. Um diese herum wurden weitere 29 Adressen schrittweise nach deren euklidischen Distanz zur gezogenen Zieladresse ermittelt (automatisiertes Verfahren durch Verwendung der Koordinaten der jeweiligen Gebäude).

Diese Gruppen von jeweils 30 Adressen wurden sodann den Einwohner:innenämter der entsprechenden Gemeinden mit der Bitte zugestellt, diese schrittweise mit 30 Haushalten aufzufüllen. Dabei ist es oftmals der Fall, dass lediglich einige wenige der 30 Zieladressen jeder Gruppe benötigt wurden, da es beispielsweise möglich ist, dass das gezogene Ziel ein Mehrfamilienhaus/Wohnblock mit mehreren Haushalten ist. Übersteigt die Anzahl Haushalte an einer bestimmten Adresse die Zahl von 30 Einheiten, wurden diese zufällig ausgewählt. Die entsprechenden Kontaktdaten wurden uns sodann von den Einwohner:innenämter für die Kontaktaufnahme zur Verfügung gestellt.

## **Wie werden die Kontaktdaten verwendet und der Datenschutz garantiert?**

Die Kontaktdaten wie auch die gesammelten Umfragedaten werden auf einem physischen Serverlaufwerk der Universität Bern gespeichert. Nur Personen mit einer Anstellung an der Universität Bern und die direkt mit der Datenerhebung beschäftigt sind (5 Personen) haben auf dieses Laufwerk Zugriff. Die Kontaktdaten werden auf diesem Serverlaufwerk zudem in einem virtuellen, verschlüsselten Laufwerk gesichert, so dass selbst bei einer physischen Entwendung des Servers die Daten weiterhin geschützt sind. Schliesslich werden die Kontaktdaten nur für die Zeit der Befragung aufbewahrt (2 Monate) und danach ohne Möglichkeit auf Wiederherstellung vernichtet.

Bei weiteren Fragen wenden Sie sich bitte an [christoph.zangger@unibe.ch](mailto:christoph.zangger@unibe.ch)