

Bevorzugter Zitierstil für diesen Vortrag

Bernard M. und M. Frick (2003) 1'000'000 Personen und ihr Verhalten: Erzeugung einer künstlichen Stichprobe für den Grossraum Zürich, IVT Seminar, ETH Zürich, Juni 2003.

1'000'000 Personen und ihr Verhalten: Erzeugung einer künstlichen Stichprobe für den Grossraum Zürich

M. Bernard, M. Frick

IVT
ETH
Zürich

Juni 2003

 *Institut für Verkehrsplanung und Transportsysteme*
Institute for Transport Planning and Systems

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Übersicht

Teil A, Einleitung

- reale Stichprobe vs. künstliche Stichprobe
- Datengrundlagen
- Ermittlung von Aktivitätsmustern
- „künstlicher“ Verkehr
- Häufigkeiten von Aktivitäten

Teil B, Erzeugung von synthetischen Populationen
mittels Iterativem Proportionalem Fitting und
Monte Carlo Sampling

Stichproben: real und künstlich

Reale Stichproben:

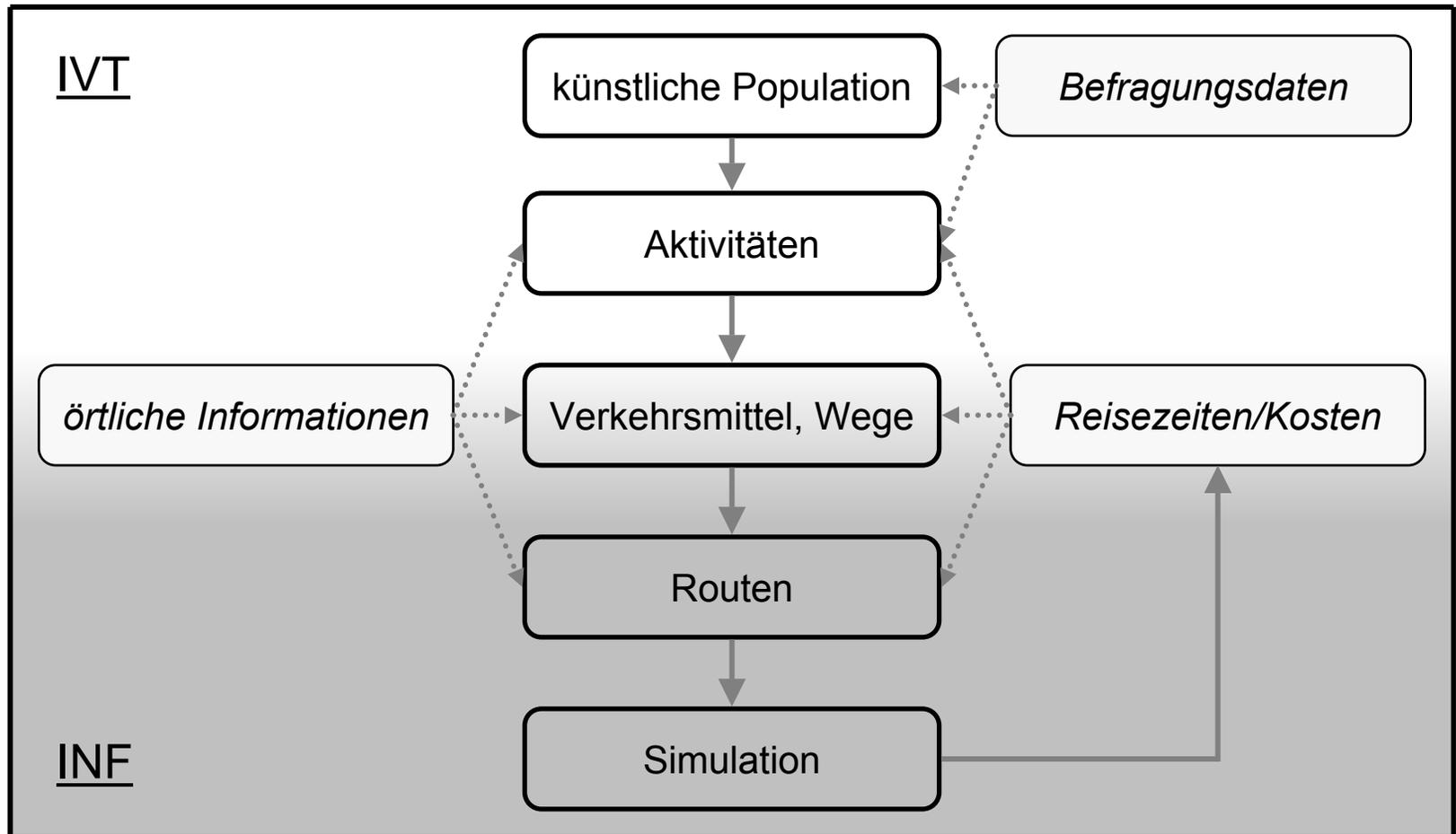
- durch Umfragen/Messungen erhoben
- i. d. R. kleiner Stichprobenumfang zur Gesamtheit
- meist nur mit hohem Aufwand zu erheben

Künstliche Stichproben, basierend auf realen Stichproben:

- erzeugte Grundgesamtheit
- modifizierte Daten für hypothetische Szenarien
- leichtere Erzeugbarkeit
- Aber: keine zusätzlichen Informationen

Ablaufschema

Übersicht der Projektteile (allgemein):
Zusammenhänge der einzelnen Komponenten



Datengrundlagen

Public Use Sample (PUS):

- basiert auf Volkszählung der Schweiz, alle 10 Jahre seit 1850
- herausgegeben vom Bundesamt für Statistik (BFS)
- zufällige 5%-Stichprobe der Personendaten der Volkszählung
- beinhaltet Schlüsseldaten über Bevölkerung, Haushalte, Arbeitsplätze und Wohnungen

Mikrozensus (MZ):

- telefonische bzw. schriftliche Umfrage zum Verkehrsverhalten
- herausgegeben vom Bundesamt für Raumentwicklung (ARE) und dem Bundesamt für Statistik (BFS)
- kleine Stichprobe vgl. mit PUS, ca. 4‰ der Bevölkerung im Jahr 2000
- beinhaltet Angaben über Personen, Haushalte, ÖV-Abonnements und Wege/Etappen für einen Tag

Ermittlung von Aktivitätsmustern (I)

Daten aus Mikrozensen (Auszug):

Struktur für jede befragte Person,
 Beispiel aus MZ 2000

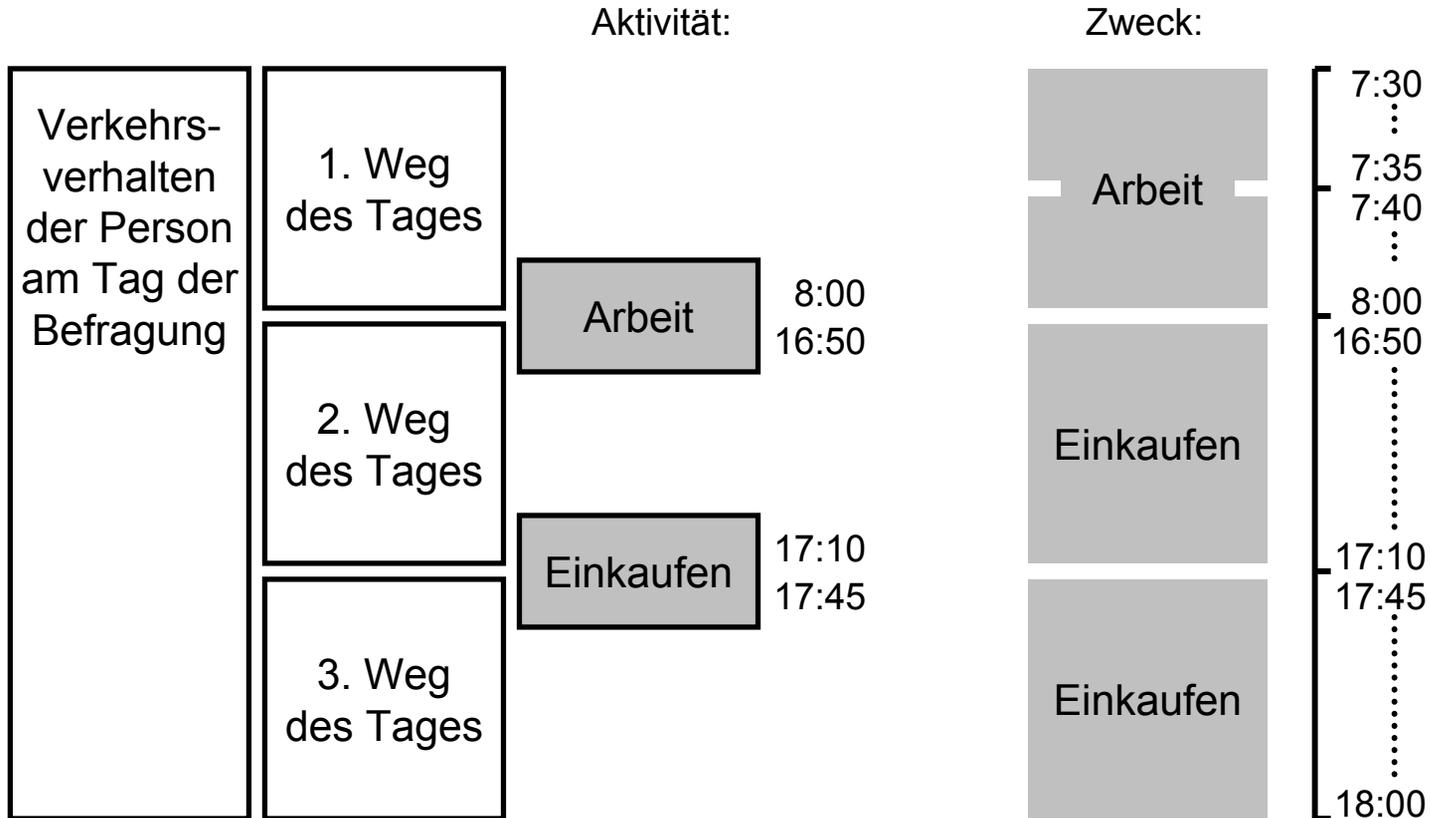
Verkehrsmittel: Zweck:

Verkehrs- verhalten der Person am Tag der Befragung	1. Weg des Tages	1. Etappe	Auto	Arbeit
		2. Etappe	Bahn	
	2. Weg des Tages	3. Etappe	Bahn	Einkaufen
		4. Etappe	Auto	Einkaufen
	3. Weg des Tages	1. Etappe	Auto	Einkaufen
		2. Etappe	Bahn	

Ermittlung von Aktivitätsmustern (II)

Daten aus Mikrozensen (Auszug):

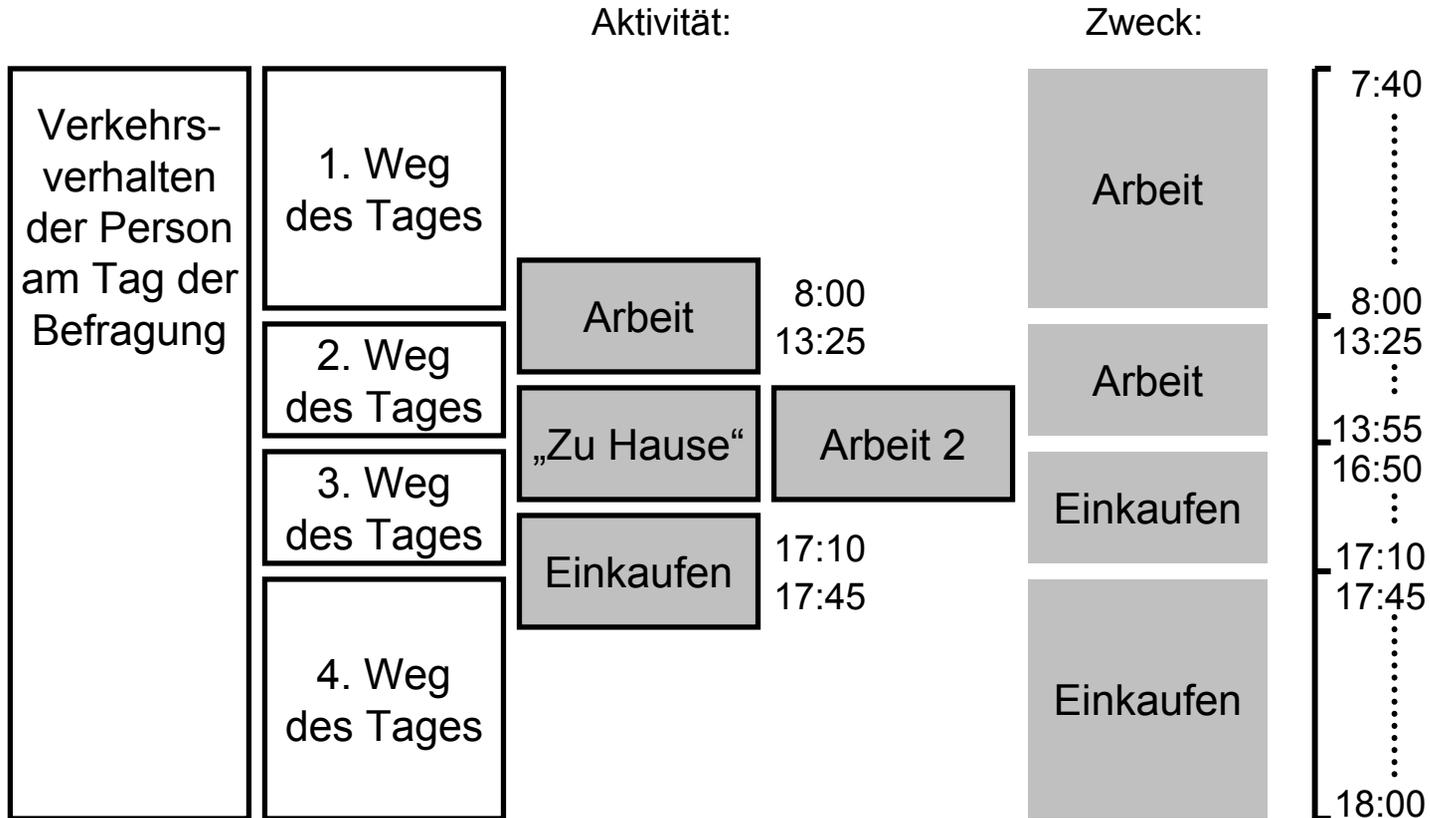
Ermitteln von Aktivitäten, Beispiel 1



Ermittlung von Aktivitätsmustern (III)

Daten aus Mikrozensen (Auszug):

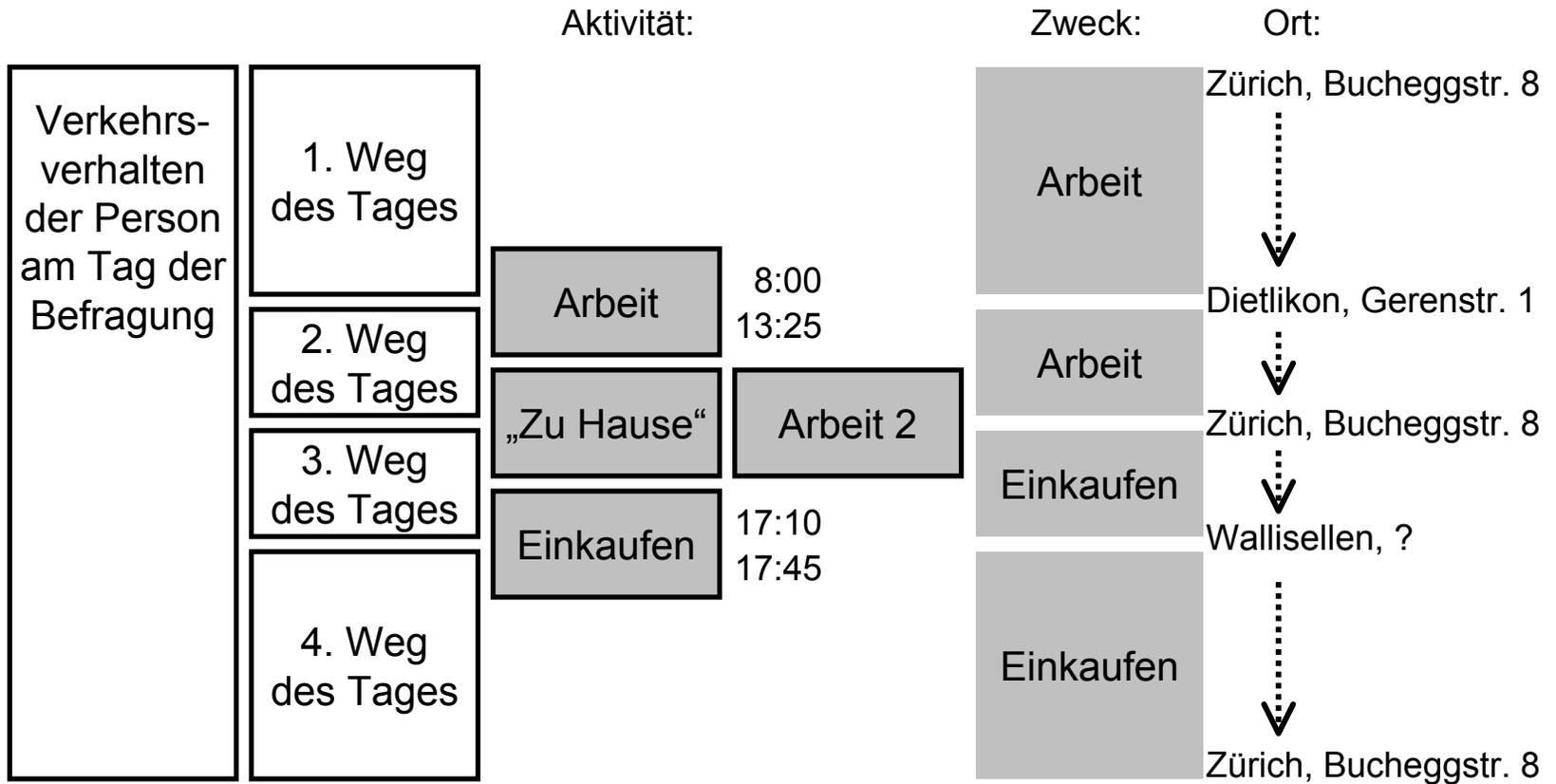
Ermitteln von Aktivitäten, Beispiel 2



Ermittlung von Aktivitätsmustern (IV)

Daten aus Mikrozensen (Auszug):

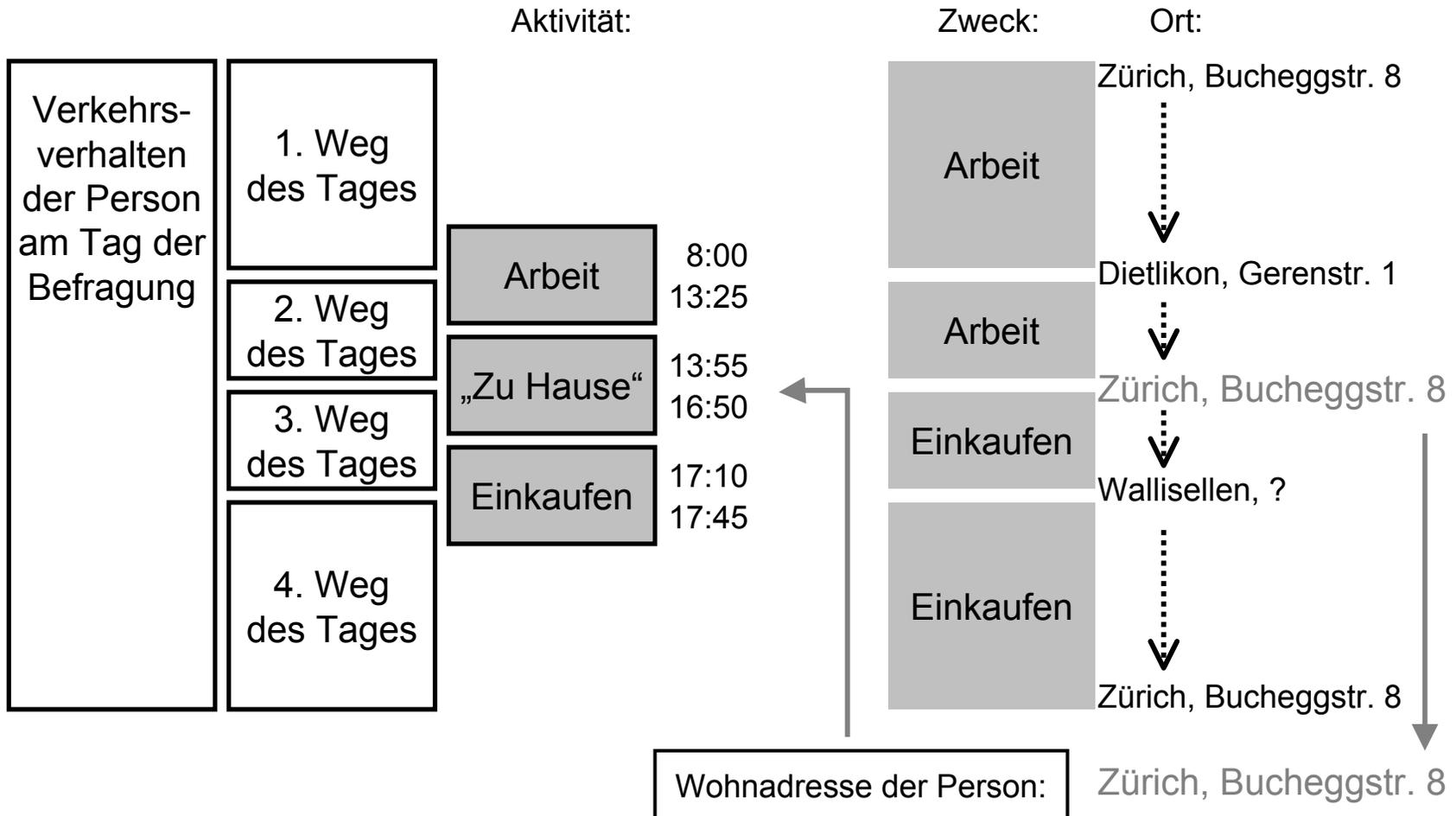
Ermitteln von Aktivitäten, Beispiel 2



Ermittlung von Aktivitätsmustern (V)

Daten aus Mikrozensen (Auszug):

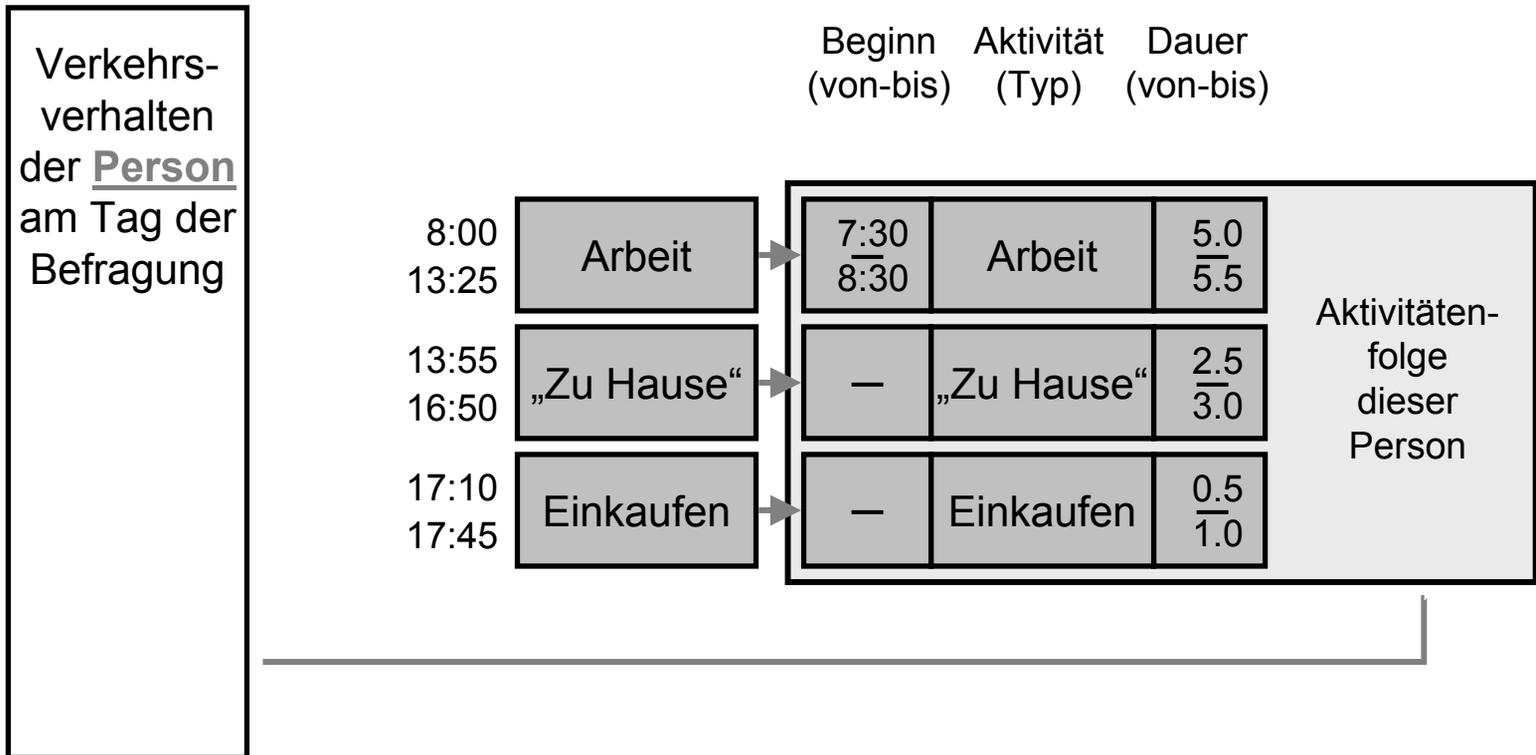
Ermitteln von Aktivitäten, Beispiel 2



Ermittlung von Aktivitätsmustern (VI)

Daten aus Mikrozensen (Auszug):

Ermittlung von Aktivitätenfolgen



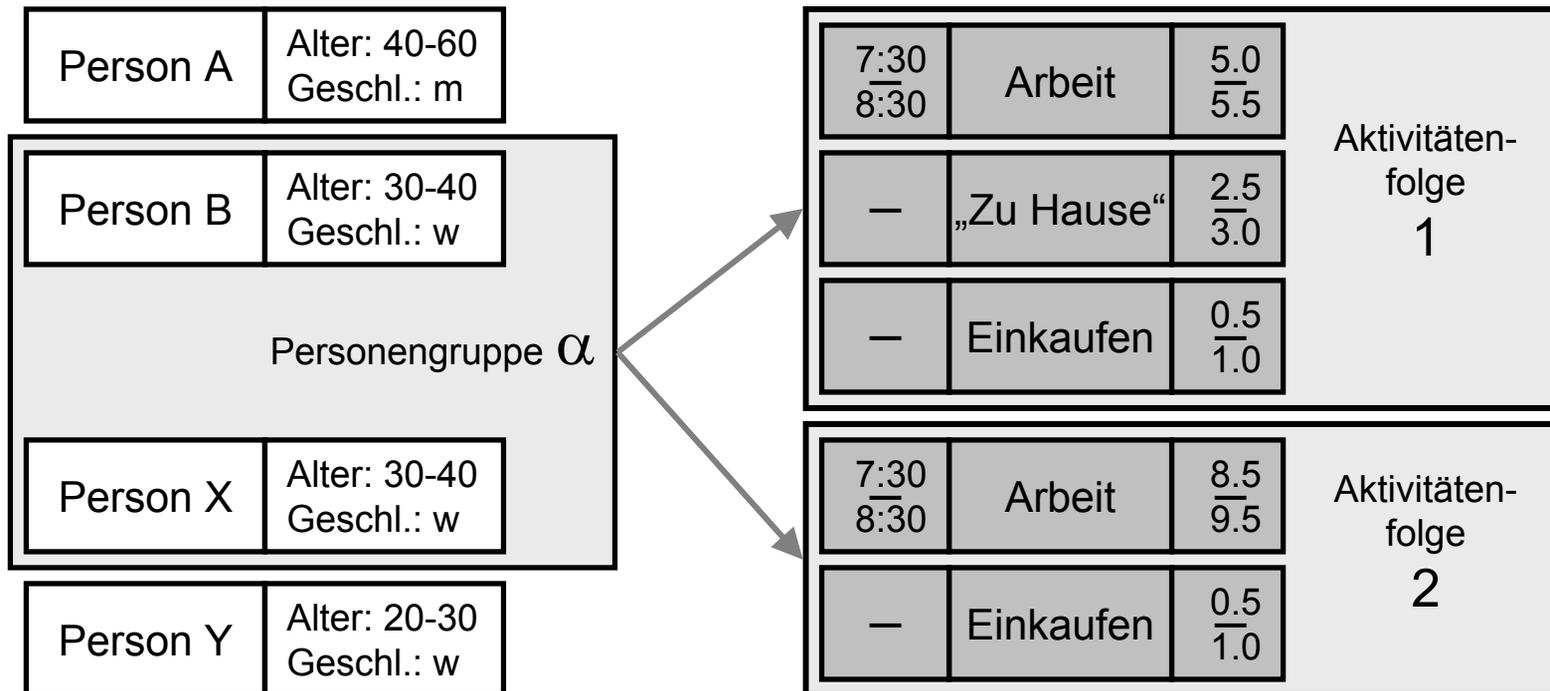
Ermittlung von Aktivitätsmustern (VII)

Daten aus Mikrozensen (Auszug):

Relation zwischen Personengruppen
und Aktivitätenfolgen

Aktivitätengruppe:

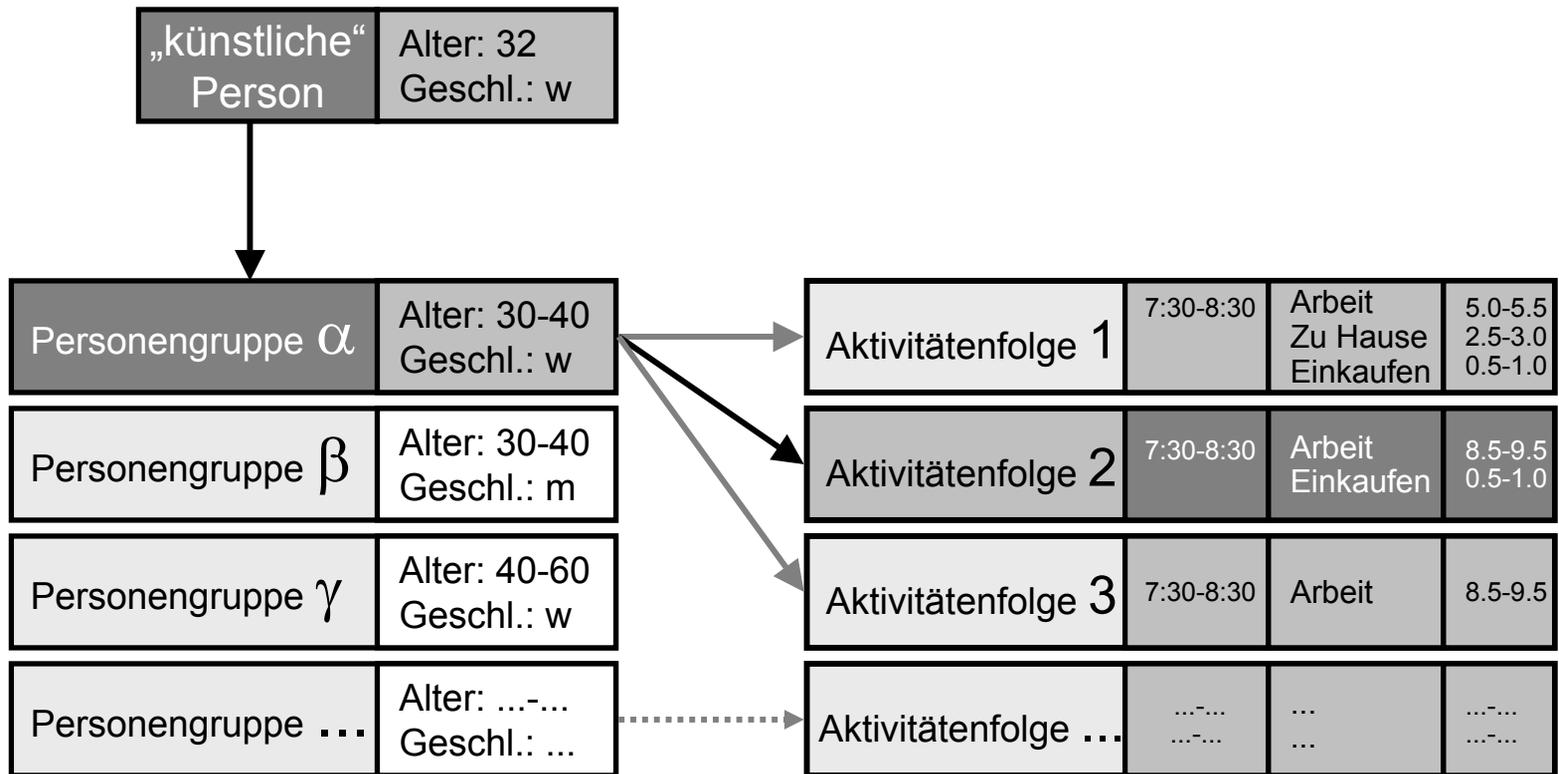
Beginn Aktivität Dauer
(von-bis) (Typ) (von-bis)



Verkehrsverhalten von „künstlichen“ Personen

Zuweisen von Aktivitätenfolgen:

Auswahl der zugehörigen Aktivitätenfolge durch zufällige Monte-Carlo-Ziehung für die Personengruppe



Ausgewählte Häufigkeiten

Häufigkeiten für Zuweisung von Aktivitätenfolgen:

Datengrundlage: Mikrozensus der Jahre 1989, 1994 und 2000, Umfang 65'617 Personen

Kategorien für Personengruppen (Beispiele):

- Geschlecht (3: m, w, k. A.)
- Alter (6: -9, -17, -29, -39, -59, 60-..)
- PW-Besitz (2: ja, nein)
- ÖV-Abo (2: ja, nein)
- Wohngemeinde (~3000)

Personengruppen mit: 30-40 Jahren, weiblich	Anzahl: 6'120 Anteil: 9.33%
Personengruppen mit: 30-40 Jahren, männlich	Anzahl: 6'282 Anteil: 9.57%
Personengruppen mit: 40-60 Jahren, weiblich	Anzahl: 10'122 Anteil: 15.43%
Personengruppen mit: 40-60 Jahren, männlich	Anzahl: 10'572 Anteil: 16.11%

Aktivitätenfolge:	Anteil:
Arbeit	13.18%
Freizeit	8.65%
Einkauf	7.28%
Arbeit - zu Hause - Arbeit	2.97%
Freizeit - Freizeit	1.89%
Arbeit - Freizeit	1.88%
Arbeit - zu Hause - Freizeit	1.84%
Arbeit - Einkauf	1.47%
...	
„Rundfahrt“	0.05%

IPF Übersicht

- Mikrosimulation in der Verkehrsplanung
- Generierung synthetischer Populationen (Idee)
- Verfügbare Datenquellen in der Schweiz
- Räumliche Disaggregation von soziodemographischen Daten
- Algorithmus zur Erzeugung hochdimensionaler Häufigkeitstabellen
- Zwei Schritt IPF-Methode und Ergebnisse für den Kanton Zürich

Mikrosimulation, Überblick

Ziel:

Entwicklung eines Werkzeuges zur Vorhersage von Langzeitverkehrsverhalten und Landnutzung auf einem mikroskopischen Niveau

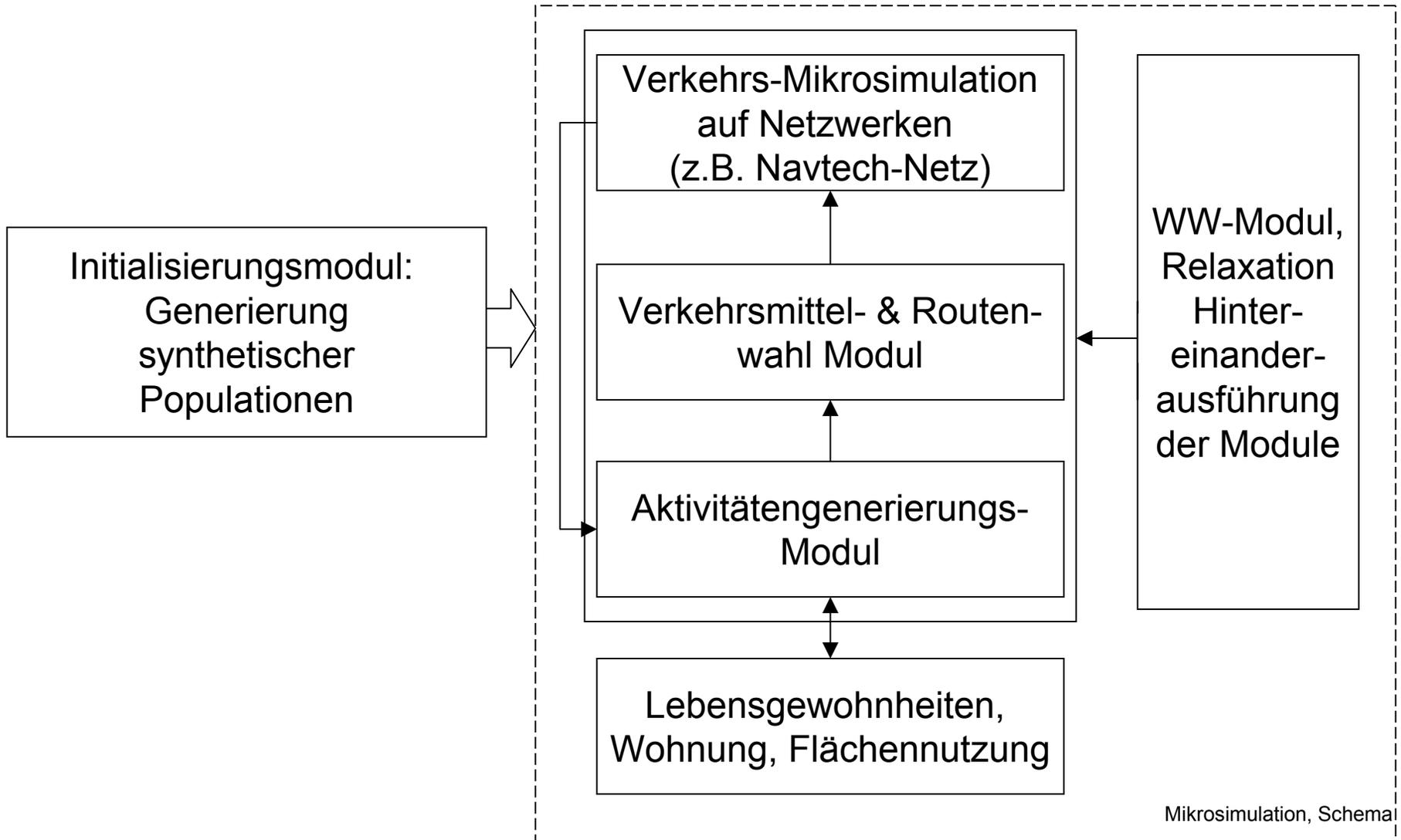
Grundlage:

Empirisch erhobene Daten von Personen, ihrer soziodemographischen Variablen und Aktivitätenmuster (von Umfragen)

Ergebnis:

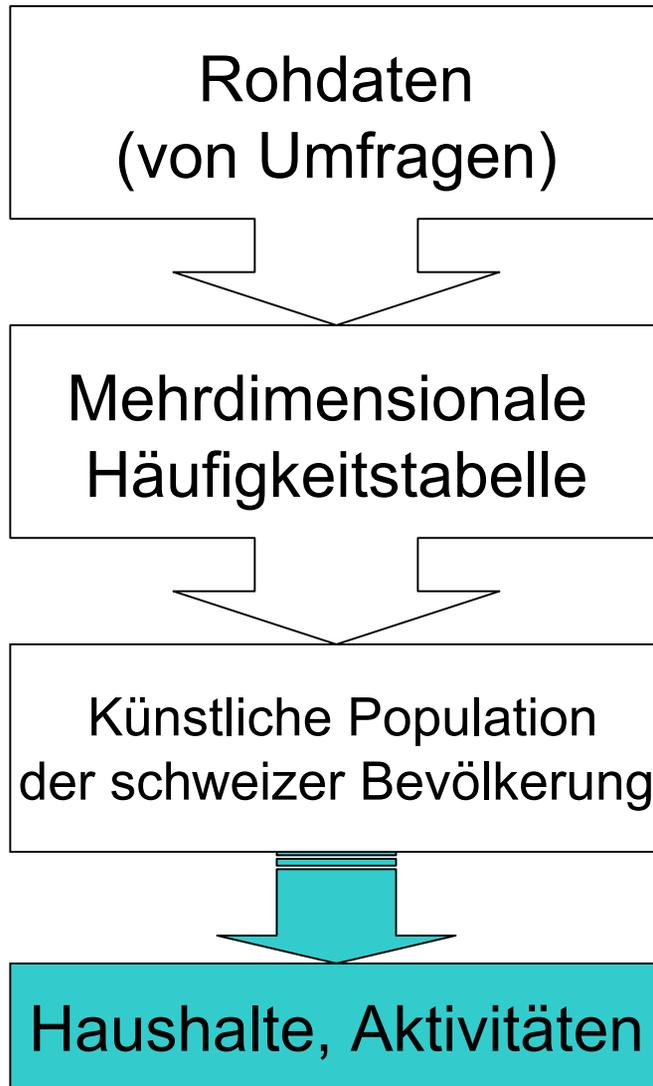
Reisezeitmatrix für individuelle Agenten und ihrer Aktivitäten (Widerstandmatrix auf mikroskopischer Ebene)

Mikrosimulation, Module



Mikrosimulation, Schema

Erzeugen künstlicher Populationen, Idee



- Verkehrsmicrozensus (0.4%)
- PUS Schweiz (5%)
- Volkszählung (aggregiert)
- Hektardatenbank

Mögliche soziodemographische Variablen:

- Alter (**in Jahren**)
- Geschlecht (**männlich / weiblich**)
- Wohnort (**xy-Koordinaten Hektare**)
- Beruf (**Kategorien**)
- Einkommen (**in SFR pro Jahr**)
- Haushaltsgrösse (**Anzahl der Personen**)
- Autobesitz (**ja / nein**)
- Auto Verfügbarkeit pro Haushalt (**Anz. Autos**)
- Zeitkartenbesitz (**Halbtax -, Monatskarte, GA**)

Gewünschtes Ergebnis:

P 1 (31, m, (6400,2345), 5, 80000, 3, ja, 2, nein)

...

P n (54, f, (5327,3476), 1, 57000, 4, ja, 2, Halbtax)
n \approx 7 Millionen (**Bevölkerung Schweiz**)

Datenquellen

Ziel:

- Synchronisation der verschiedenen Datenquellen die für die Öffentlichkeit zugänglich sind. Diese Daten liefern die Basis für das Erzeugen der Häufigkeitstabelle für die gesamte Population
- Disaggregation der Daten
 - zu individuellen Agenten
 - mit einer möglichst feinen räumlichen Auflösung (Gemeinden, Hektar Raster, xy-Koordinaten)

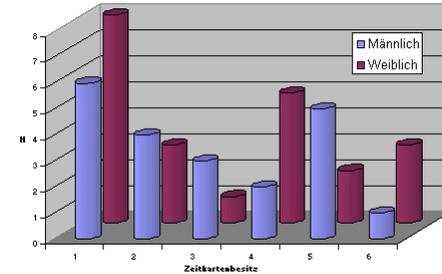
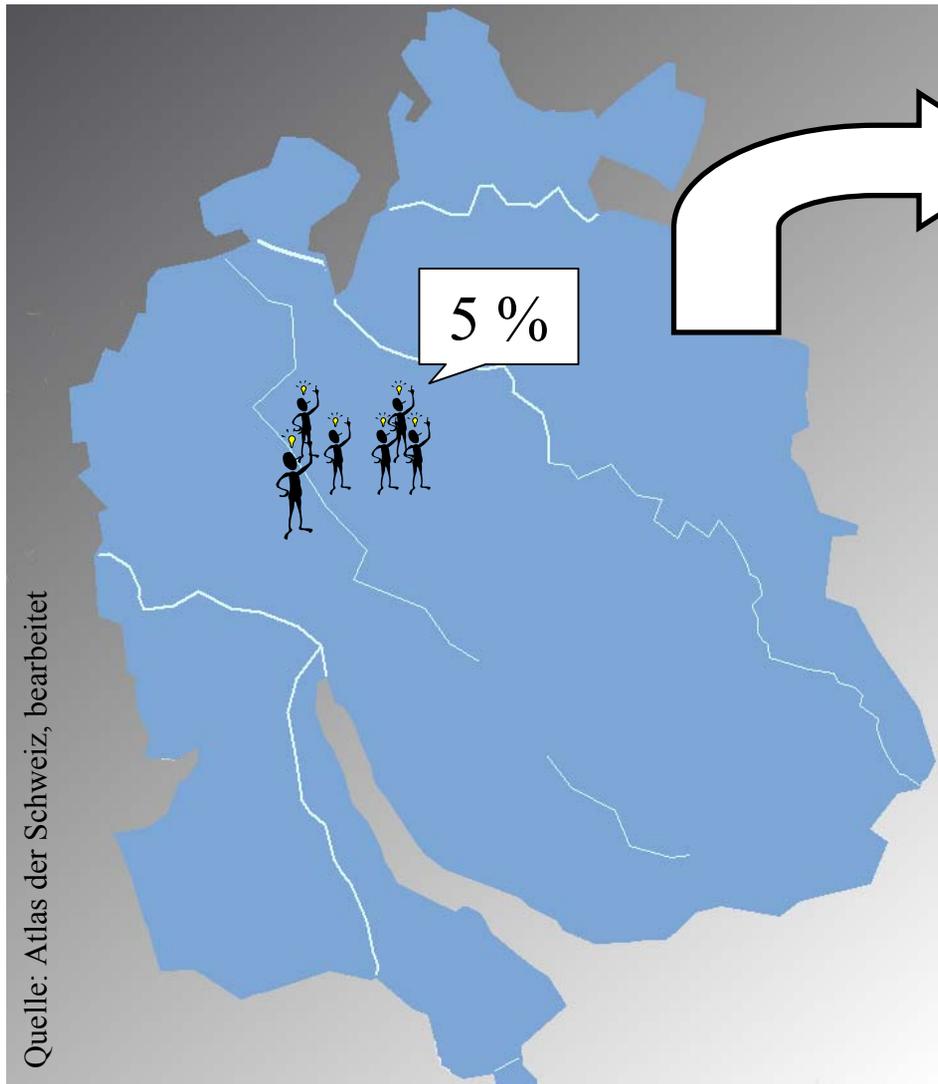
Datenquellen, einige Probleme

Umfragen sind heterogen:

- Verschiedene Fragestellungen
- Verschiedene Auflösung der soziodemographischen Variablen
- Verschiedenes Aggregationsniveau der Variablen
- Fehlende Einträge (Keine Angaben, Datenschutz)
- Verschiedene räumliche Auflösung der Daten
- **Die Korrelationsstruktur** hat eine räumliche Variation

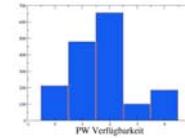
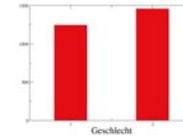
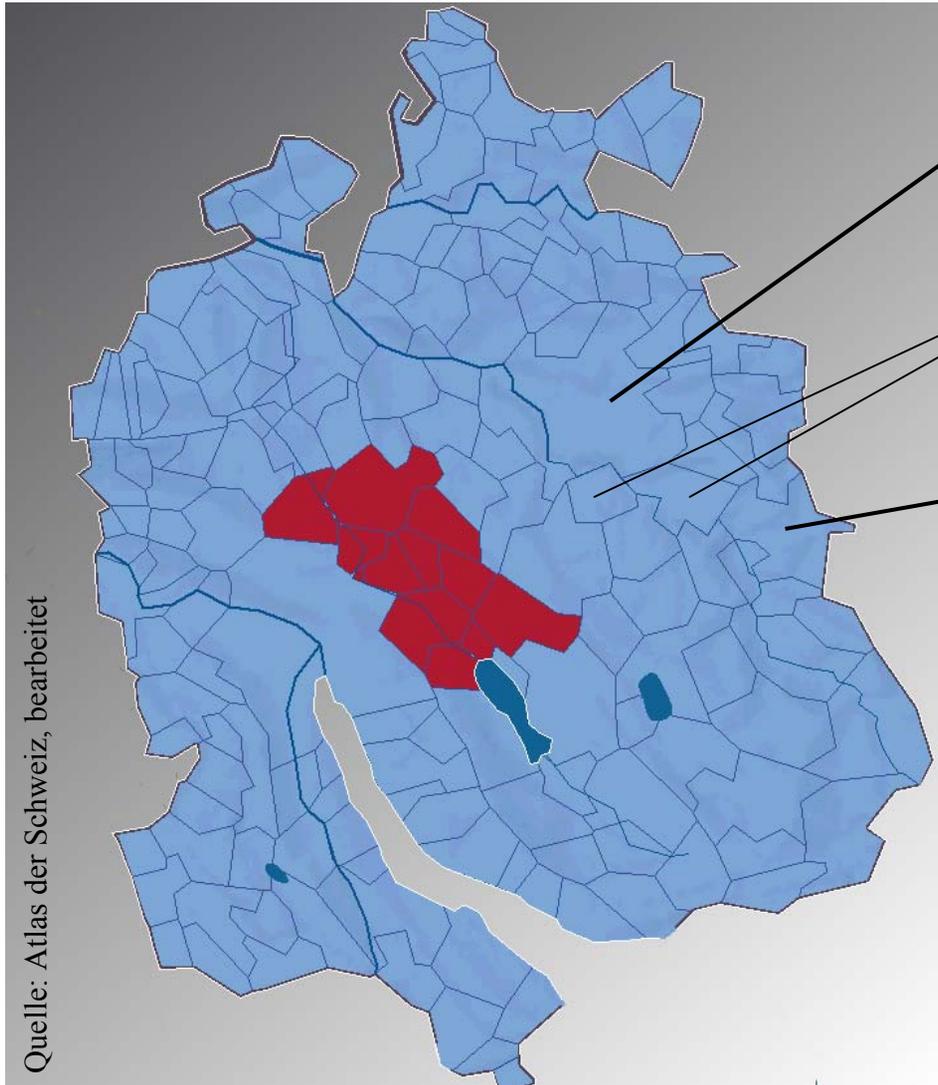
Ist die erhobene Stichprobe repräsentativ ?

Räumliche Disaggregation, Kanton Zürich



- Für jede Person kennt man die soziodemographischen Werte
- ⇒ Konstruktion einer hochdimensionalen Häufigkeitstabelle für die 5% Stichprobe ist möglich (Anfangsverteilung)

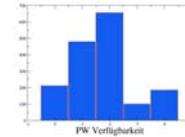
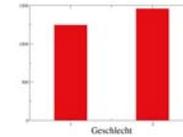
Räumliche Disaggregation, Gemeinden



Variable1, Variable2, ...

⋮
⋮

⋮
⋮



Variable1, Variable2, ...

Auf der Gemeinde Ebene existieren nur Informationen über die Randverteilungen für jede soziodemographische Variable

Iteratives Proportionales Fitting (IPF), 3D Formeln

Häufigkeitstabelle

$$\pi_{ijk} \in R^+, i \in \{1, \dots, I\}, j \in \{1, \dots, J\}, k \in \{1, \dots, K\}$$

Zwangsbedingungen
entsprechend der Rand-
Verteilungen

$$n\pi_{ij\bullet} = x_{ij\bullet}, \quad n\pi_{i\bullet k} = x_{i\bullet k}, \quad n\pi_{\bullet jk} = x_{\bullet jk},$$

$$n := n\pi_{\dots} = x_{\dots}$$

Schritt 1

$$\pi_{ijk}^{(1)} = \frac{1}{n} \frac{x_{ij\bullet} \pi_{ijk}^{(0)}}{\pi_{ij\bullet}^{(0)}}$$

Schritt 2

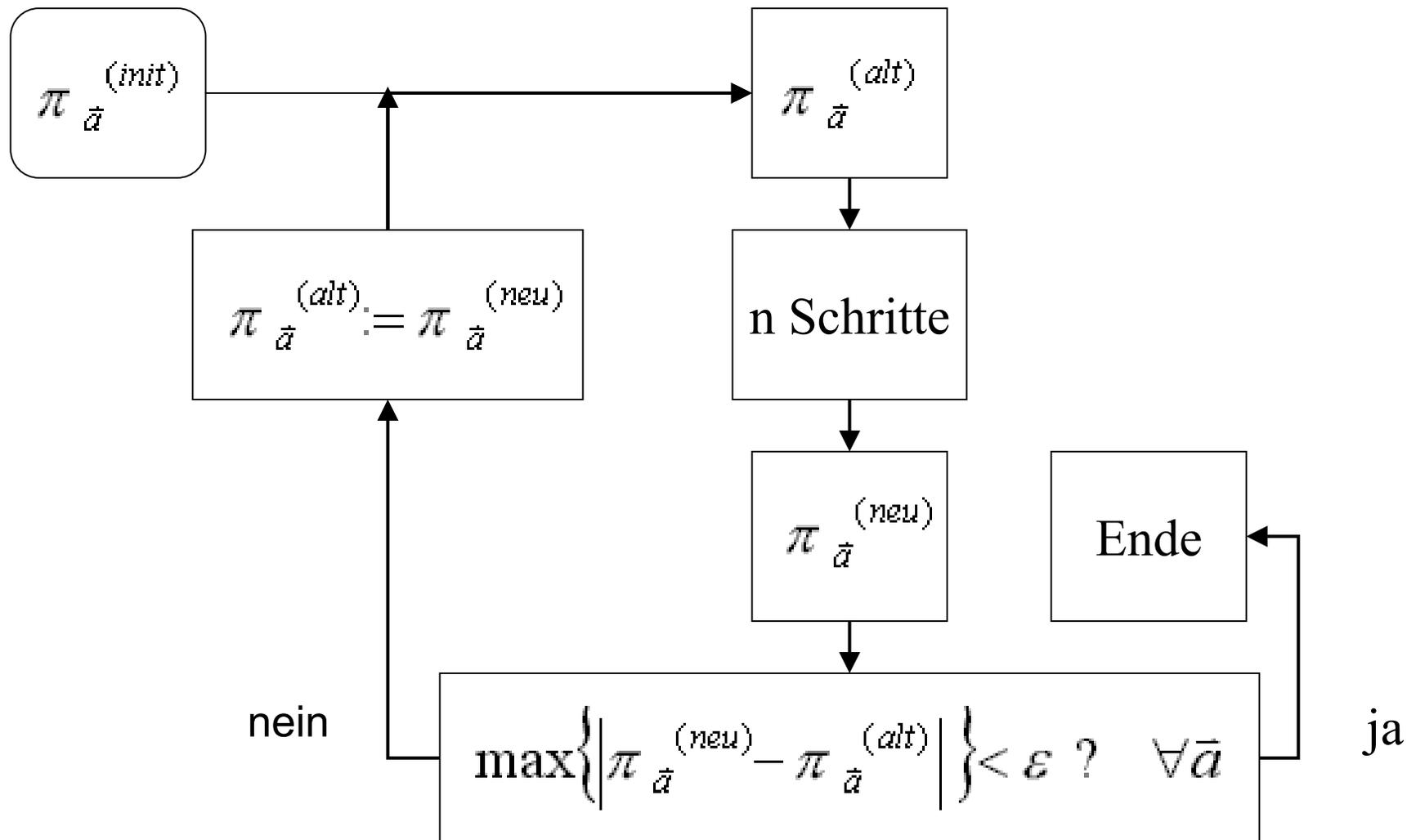
$$\pi_{ijk}^{(2)} = \frac{1}{n} \frac{x_{i\bullet k} \pi_{ijk}^{(1)}}{\pi_{i\bullet k}^{(1)}}$$

Schritt 3

$$\pi_{ijk}^{(3)} = \frac{1}{n} \frac{x_{\bullet jk} \pi_{ijk}^{(2)}}{\pi_{\bullet jk}^{(2)}}$$

eine Iteration

Iteration, Schema



Erster IPF Schritt, Kanton Zürich

Altersverteilung = (1,0) =

(11593, 51128, 64662, 64225, 64684, 77601, 93278, 110767,
111411, 96505, 86962, 86131, 79676, 61763, 53412, 46781, 38128,
24765, 15942, 6984, 1403, 103, 2)

Geschlechtsverteilung = (0,1) = (613038, 634868)

IPF Schritt 1: (1,0), (0,1) und (1,1) für die 5% Stichprobe
=> (1,1) geschätzt.

Dies führt auf...

IPF Schritt 1, Kanton Zürich, Ergebnis

Alter x Geschlecht Verteilung,

$$(1,1) = \begin{pmatrix} 6745 & 4848 \\ 29882 & 21246 \\ 37286 & 27376 \\ 36268 & 27957 \\ 35773 & 28911 \\ 42132 & 35469 \\ 49533 & 43745 \\ 57257 & 53510 \\ 55522 & 55888 \\ 46194 & 50310 \\ 40248 & 46713 \\ 38771 & 47360 \\ 34865 & 44811 \\ 26265 & 35498 \\ 22260 & 31151 \\ 19171 & 27610 \\ 15390 & 22738 \\ 9858 & 14907 \\ 6290 & 9653 \\ 2737 & 4246 \\ 548 & 855 \\ 40 & 63 \\ 1 & 1 \end{pmatrix}$$

Zweiter Schritt des IPF Verfahrens in 3 Dimensionen

Wir haben $(1,1)$ vom ersten Schritt: $(1,1) = (1,1,0)$.

Wir benötigen

Alter Geschlecht Gemeinde Nr.

Alter x Gemeinde# Verteilung: $(1,0,1)$

Geschlecht x Gemeinde# Verteilung: $(0,1,1)$

als Randverteilungen um den zweiten IPF Schritt auszuführen:

$(1,0,1)$; $(0,1,1)$; und $(1,1,0)$ \Rightarrow $(1,1,1)$

- $(1,1,1)$ ist die gesuchte Alter x Geschlecht x Gemeinde Nr. Verteilung
- $(1,1,1)$ kann aufgeteilt werden in $(1,1)$ Tabellen für jede Gemeinde

Ergebnis für die Gemeinde Aeugst am Albis

Alter x Geschlecht	Alter (Ist)	Alter (Soll)	Delta
10 7	17	17	0
40 27	67	67	0
63 41	104	104	0
65 44	109	109	0
43 35	78	78	0
39 36	75	75	0
28 30	58	58	0
55 53	108	108	0
66 66	132	132	0
94 90	184	184	0
71 76	147	147	0
55 66	121	121	0
47 62	109	109	0
31 44	75	75	0
24 35	59	59	0
17 27	44	44	0
13 22	35	35	0
4 6	10	10	0
4 6	10	10	0
1 1	2	2	0
0 0	0	0	0
0 0	0	0	0
0 0	0	0	0

Volkszählungsdaten, Schweiz 1990

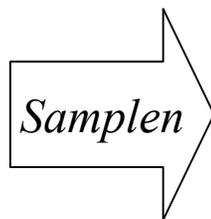
Wir haben:

- 4D-Häufigkeitstabelle:
Hektare x Erwerbstätig x Geschlecht x Alter
mit: $367\,767 \times 2 \times 2 \times 19$ (ca. 28 Mio.) verschiedenen Kategorien
(Grundlage: Geostat 1990)
- Samplen von dieser 4D Verteilung entsprechend der Häufigkeit des Auftretens jeder Kategorie führt zu 4D Tupeln mit Kategorien welche den Agenten charakterisieren.
- Nimmt man eine Gleichverteilung innerhalb der Kategorien an, kann man mit Hilfe eines einfachen Monte Carlo Samplings disaggregierte 4D Tupel bekommen. Diese beschreiben jeden Agenten auf mikroskopischer Ebene durch seine soziodemographischen Variablen.

Volkszählungsdaten, Schweiz 1990, Beispiel

(xy-Koordinaten, erwerbstätig,
Geschlecht, Alter) --> Häufigkeit

...
 (4864001117, 1, 1, 7) ---> 2
 (4864001117, 1, 1, 8) ---> 3
 (4864001117, 1, 1, 9) ---> 3
 (4864001117, 1, 1, 10) ---> 1
 (4864001117, 1, 1, 11) ---> 3
 (4864001117, 1, 1, 13) ---> 1
 (4864001117, 1, 2, 8) ---> 3
 (4864001117, 1, 2, 9) ---> 5
 (4864001117, 1, 2, 10) ---> 3
 (4864001117, 1, 2, 11) ---> 1
 (4864001117, 2, 1, 1) ---> 2
 (4864001117, 2, 1, 3) ---> 4
 (4864001117, 2, 1, 5) ---> 1
 (4864001117, 2, 1, 7) ---> 1
 (4864001117, 2, 2, 2) ---> 1
 (4864001117, 2, 2, 5) ---> 1
 (4864001117, 2, 2, 6) ---> 1
 (4864001117, 2, 2, 8) ---> 1
 (4864001117, 2, 2, 9) ---> 1
 (4864001117, 2, 2, 11) ---> 1
 ...

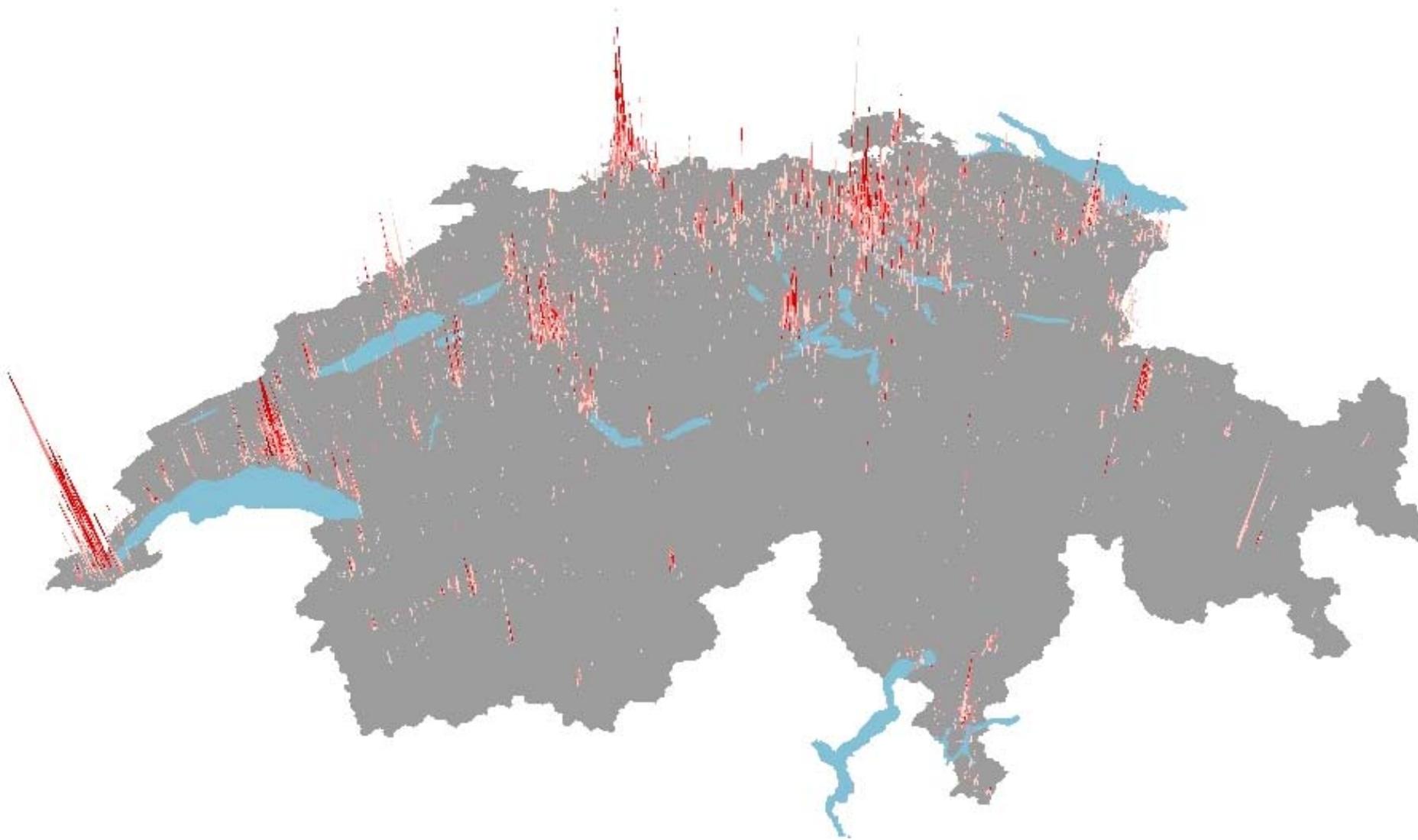


(4864001117, 1, 1, 32), (4864001117, 1, 1, 31), (4864001117, 1, 1, 36),
 (4864001117, 1, 1, 35), (4864001117, 1, 1, 37), (4864001117, 1, 1, 44),
 (4864001117, 1, 1, 43), (4864001117, 1, 1, 41), (4864001117, 1, 1, 48),
 (4864001117, 1, 1, 51), (4864001117, 1, 1, 51), (4864001117, 1, 1, 54),
 (4864001117, 1, 1, 60), (4864001117, 1, 2, 39), (4864001117, 1, 2, 37),
 (4864001117, 1, 2, 39), (4864001117, 1, 2, 41), (4864001117, 1, 2, 41),
 (4864001117, 1, 2, 43), (4864001117, 1, 2, 40), (4864001117, 1, 2, 43),
 (4864001117, 1, 2, 49), (4864001117, 1, 2, 46), (4864001117, 1, 2, 47),
 (4864001117, 2, 1, 1), (4864001117, 2, 1, 1),
 (4864001117, 2, 1, 10), (4864001117, 2, 1, 14), (4864001117, 2, 1, 11),
 (4864001117, 2, 1, 13), (4864001117, 2, 1, 21), (4864001117, 2, 1, 32),
 (4864001117, 2, 2, 5), (4864001117, 2, 2, 23), (4864001117, 2, 2, 29),
 (4864001117, 2, 2, 38), (4864001117, 2, 2, 44), (4864001117, 2, 2, 54)

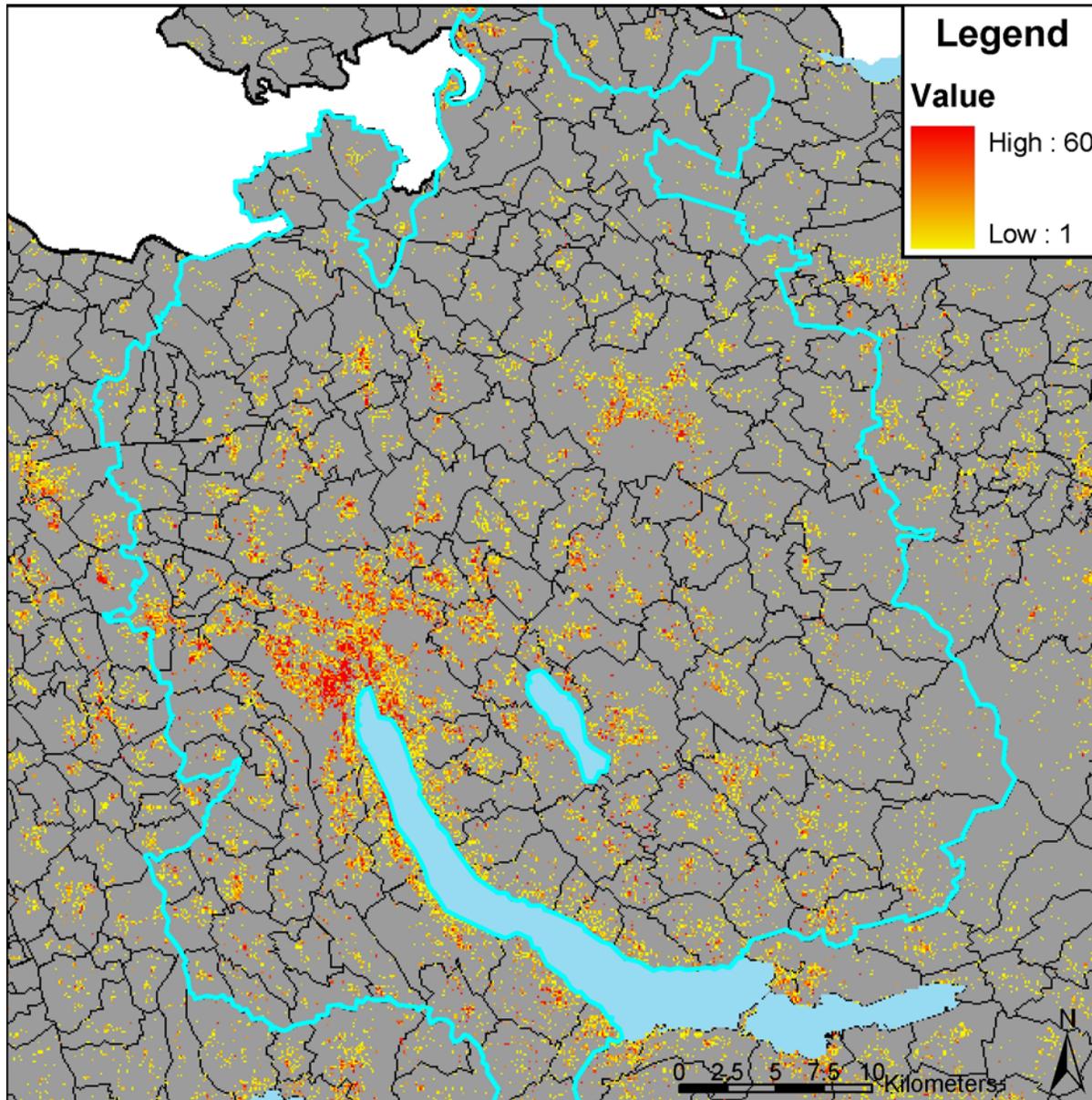
(4864001117, 1, 2, 54) z.B. ist eine Agentin, die auf Hektar x=486400, y=111700 wohnt und eine 54 Jahre alte erwerbstätige Frau repräsentiert.

 Summe 39

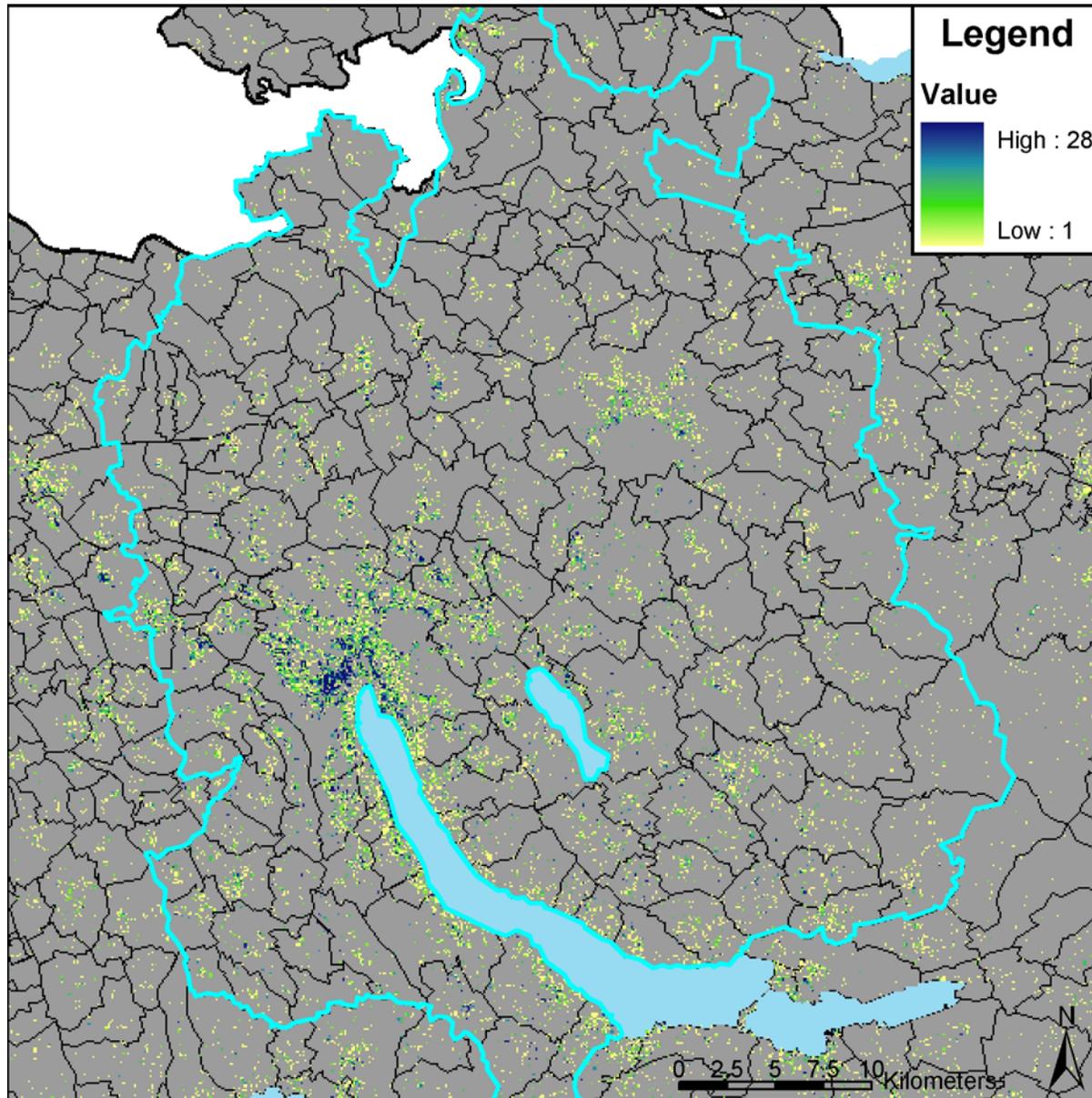
Schweiz: Berufstätige Männer zwischen 35-39



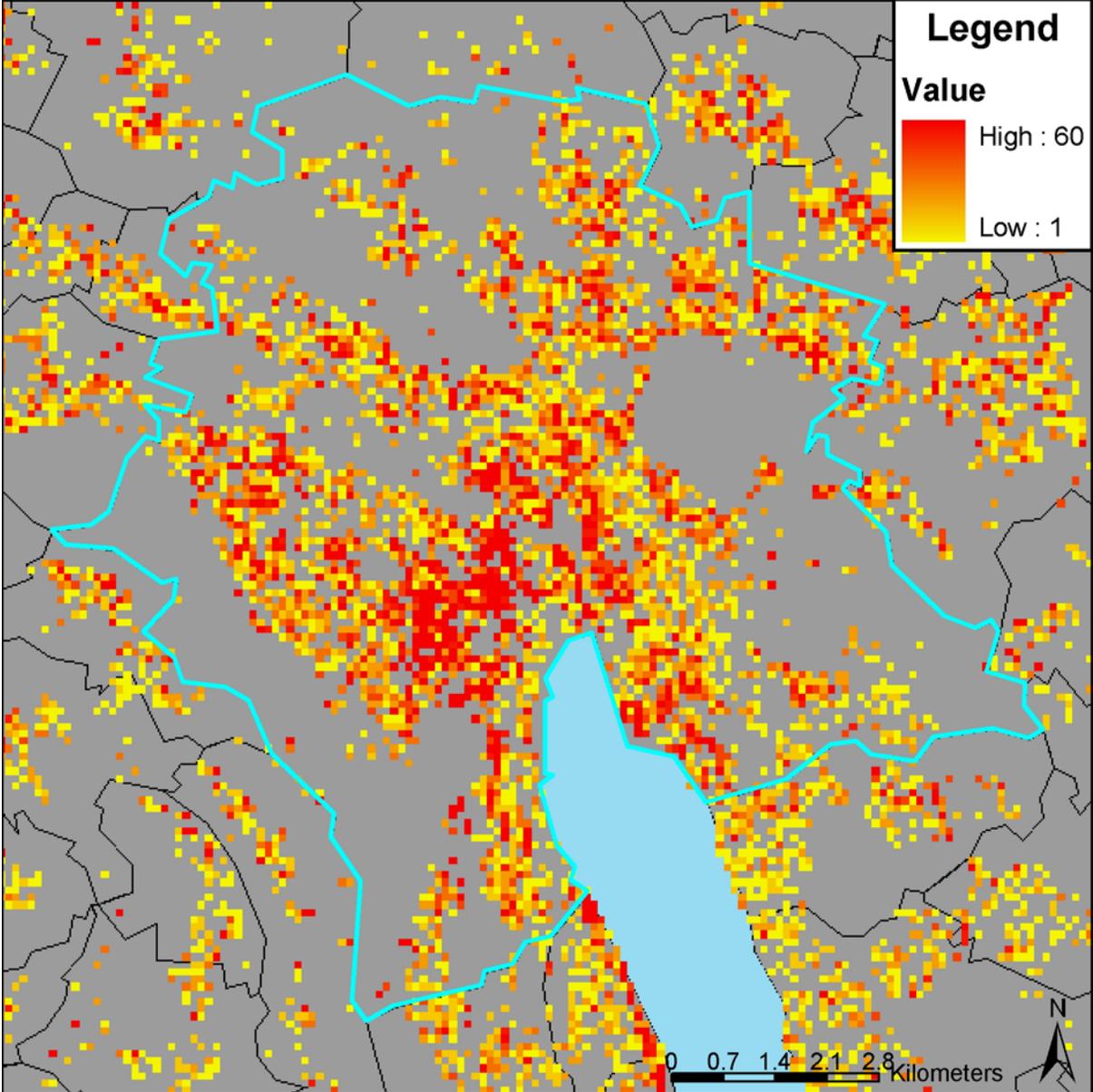
Kt. Zürich: Berufstätige Männer zwischen 35-39



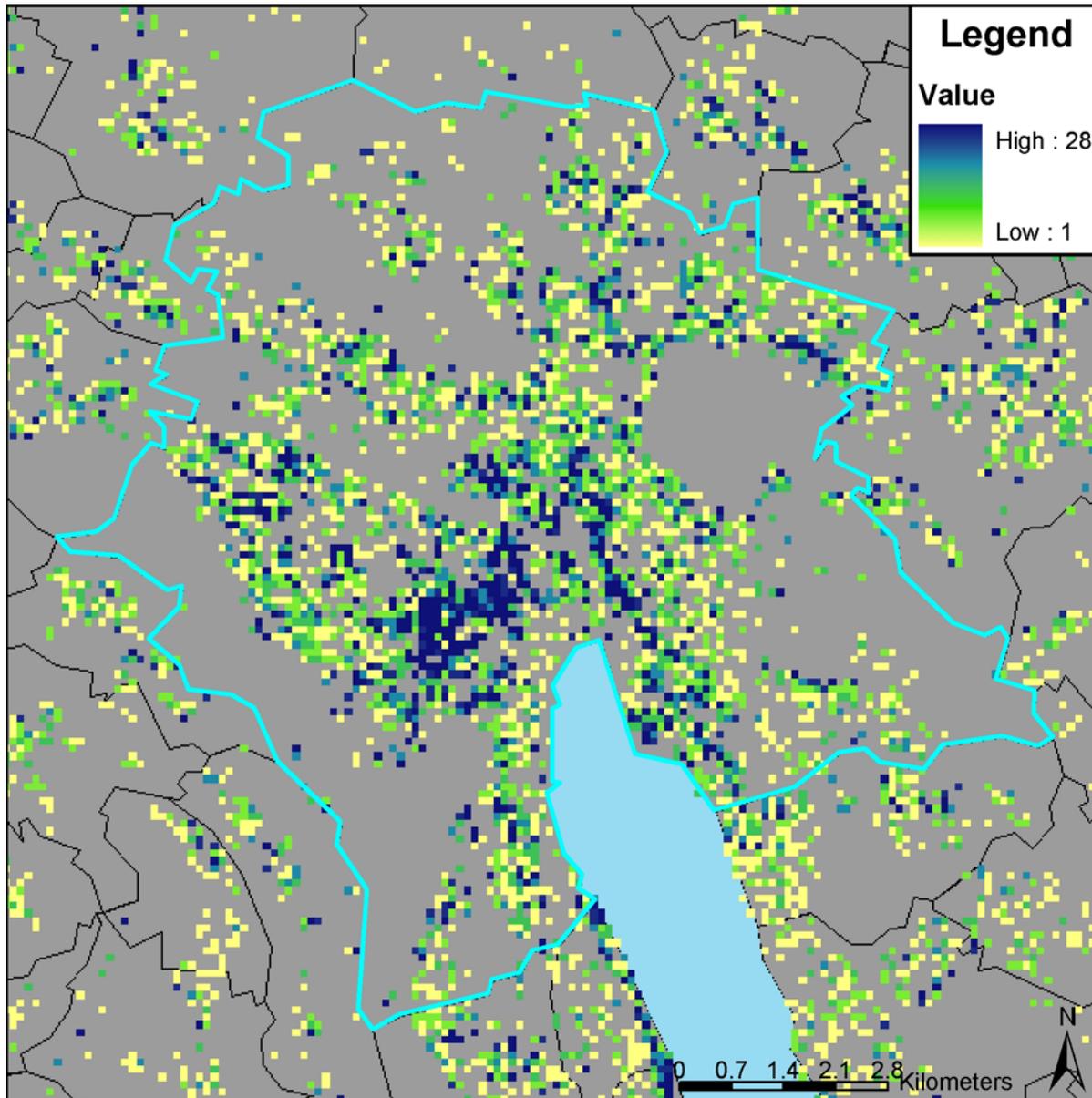
Kt. Zürich: Berufstätige Frauen zwischen 35-39



Stadt Zürich: Berufstätige Männer zwischen 35-39



Stadt Zürich: Berufstätige Frauen zwischen 35-39



Mikrozensus 2000 & VZ2000, IPF, Erklärung

IPF mit 7 Variablen für die gesamte Schweiz, (ohne Wohnort).

Start mit 1D Verteilungen und div. Anfangsverteilungen (meist 2D):

$(1,0,0,0,0,0,0)$ = Altersverteilung (23 Kategorien)

$(0,1,0,0,0,0,0)$ = Geschlecht (männlich/weiblich)

$(0,0,1,0,0,0,0)$ = Führerscheinbesitz (ja/nein); Alter > 17

$(0,0,0,1,0,0,0)$ = Verfügbarkeit PW (ja/nein); Führerscheinbesitz=ja

$(0,0,0,0,1,0,0)$ = Erwerbstätig (ja/nein); Alter > 15

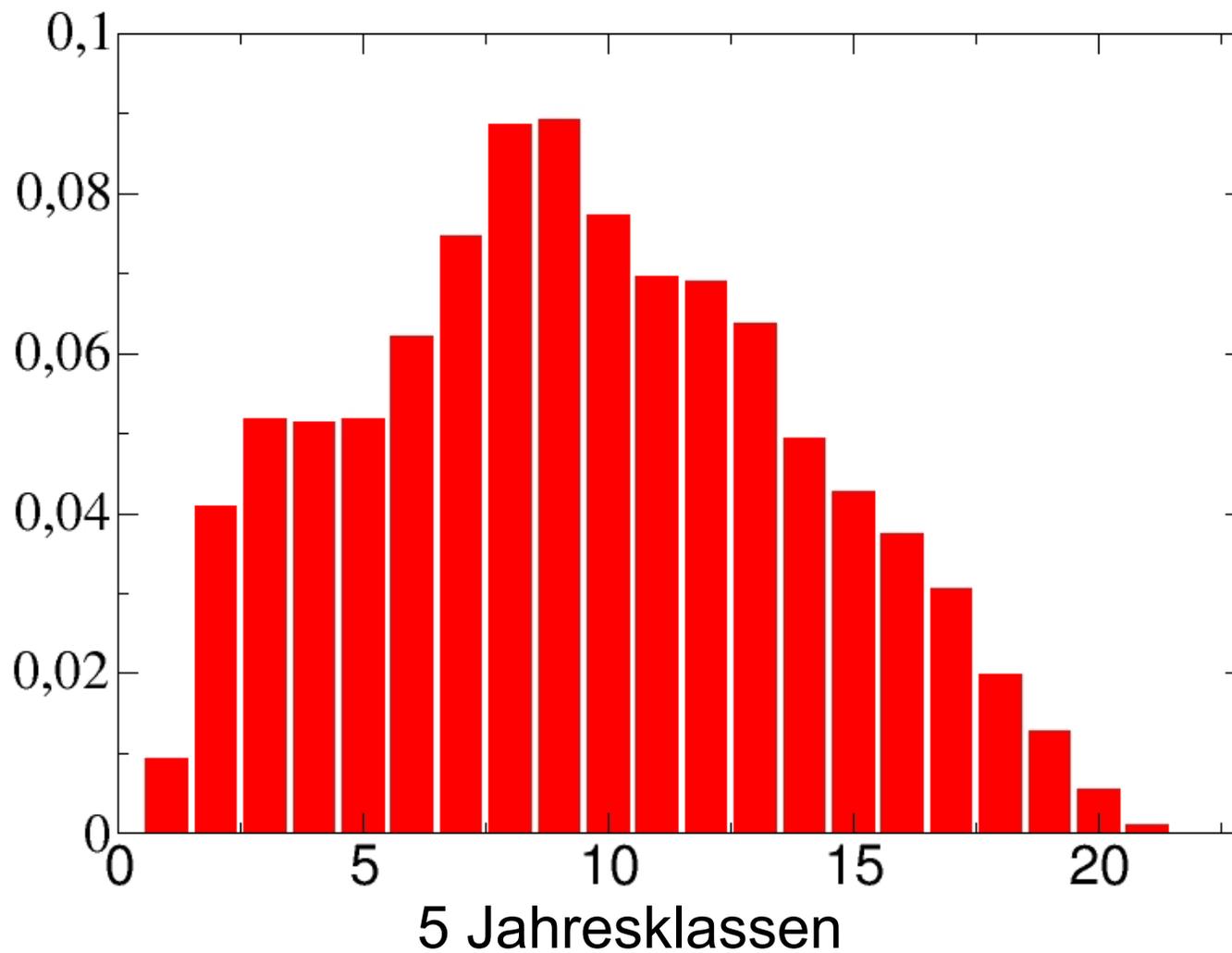
$(0,0,0,0,0,1,0)$ = Halbtaxabo (1,2,3,4); Alter > 15

$(0,0,0,0,0,0,1)$ = GA-Abo (ja/nein); Alter > 5

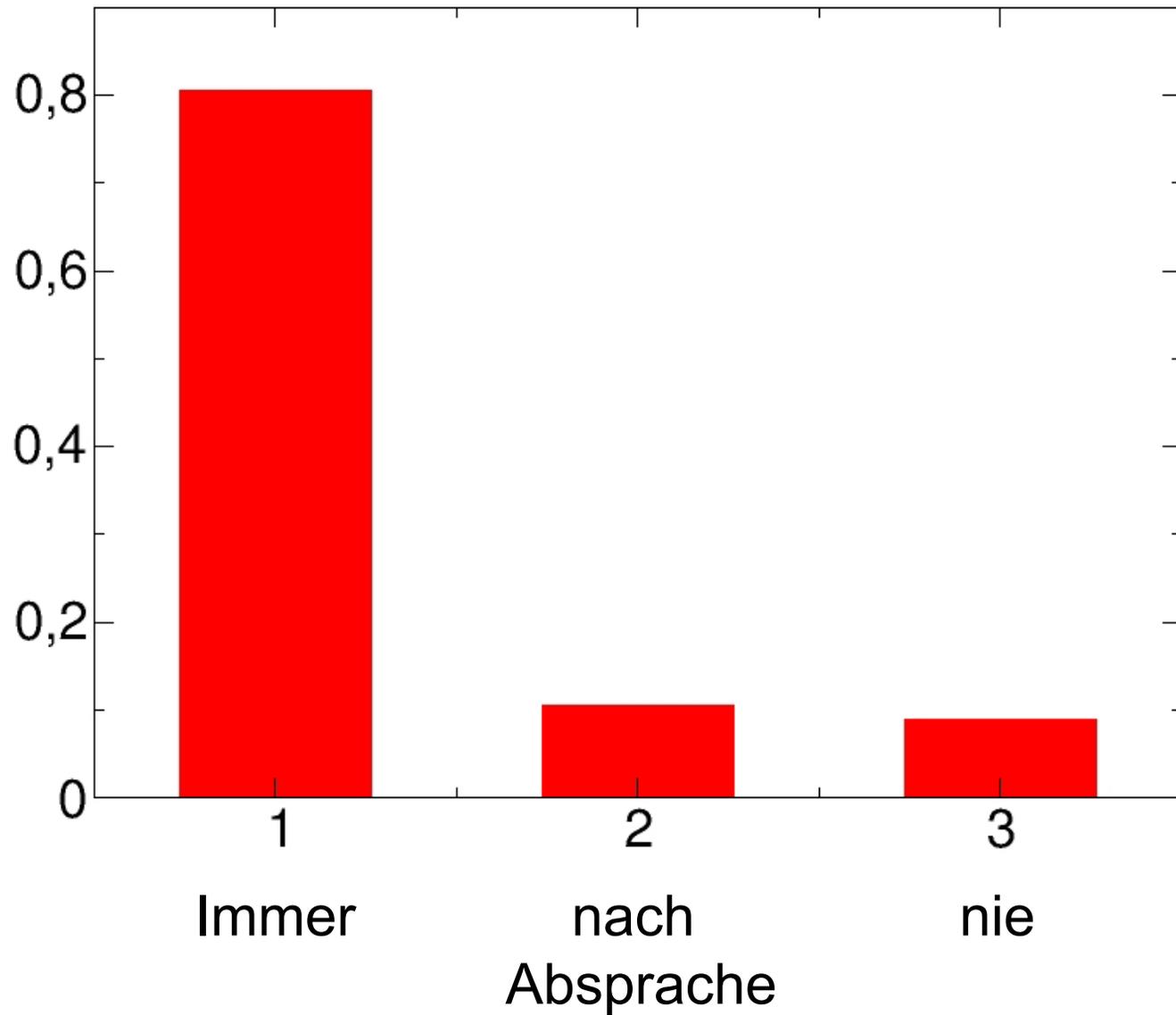
IPF nacheinander 128 mal ausgeführt, liefert 7D-Verteilung für ca. 7 Mio. Schweizer, bei 2944 möglichen Kategorien.

==> Jeder Agent ist charakterisiert durch einen 7D-Vektor.

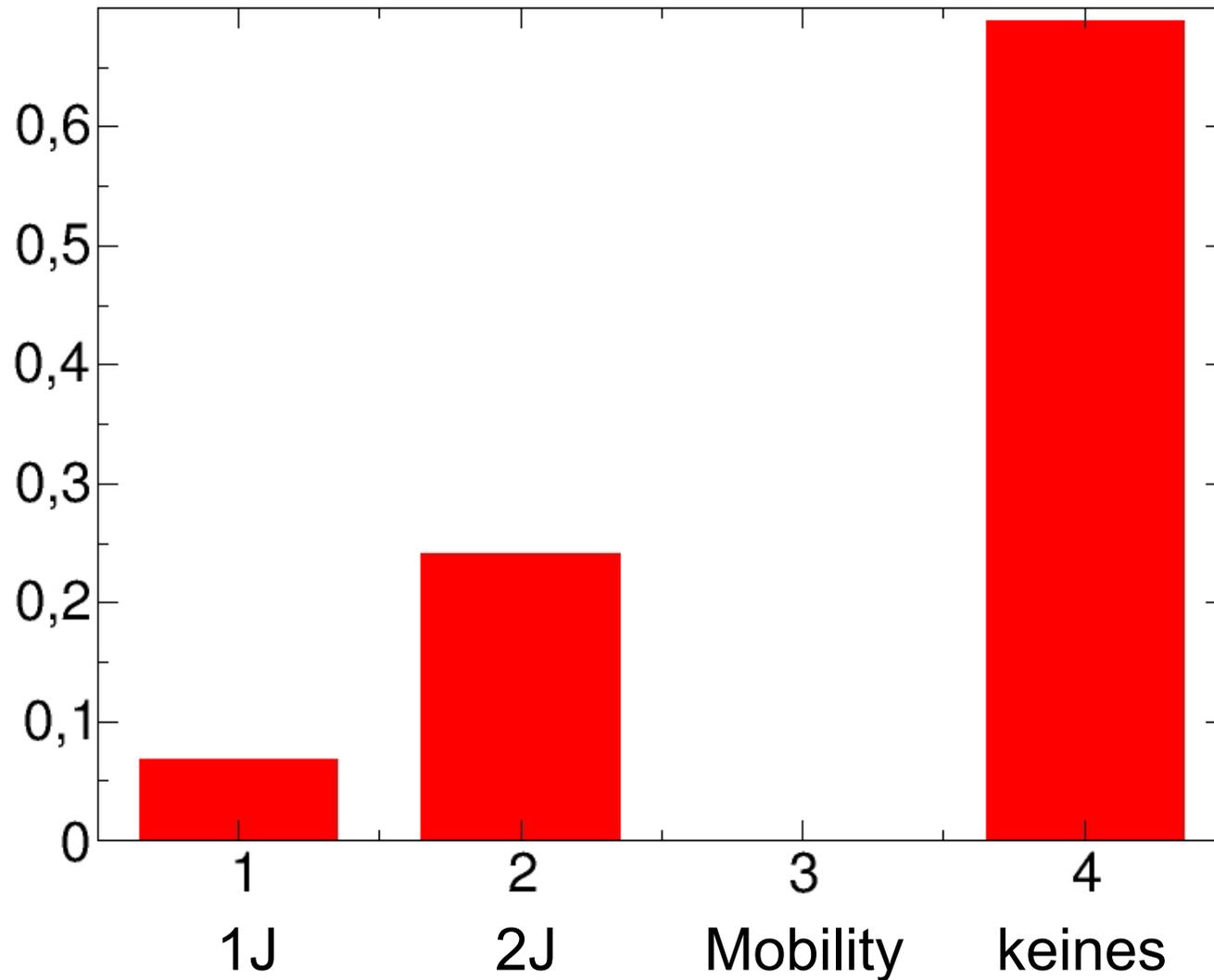
Mikrozensus 2000 & VZ2000, Altersverteilung



Mikrozensus 2000 & VZ2000, PW-Verfügbarkeit



Mikrozensus 2000 & VZ2000, Halbtaxbesitz



Mikrozensus 2000 & VZ2000, IPF, Ergebnis

...

{10, 2, 1, 2, 2, 1, 2}, {17, 2, 1, 1, 2, 4, 1}, {14, 1, 1, 1, 1, 4, 2},
{3, 2, 2, 2, 2, 4, 2}, {12, 2, 1, 1, 2, 2, 2}, {7, 1, 1, 1, 1, 4, 2},
{9, 1, 2, 2, 2, 4, 2}, {12, 2, 1, 2, 1, 2, 2}, {6, 1, 1, 1, 1, 4, 2},
{14, 2, 2, 2, 2, 4, 1}, {11, 2, 1, 1, 1, 2, 1}, {11, 1, 1, 1, 2, 4, 2},
{8, 2, 1, 2, 1, 2, 2}, {6, 2, 1, 1, 1, 1, 2}, {7, 1, 1, 1, 1, 2, 1}

...

Random
Samplen



...

{46, 2, 1, 2, 2, 1, 2}, {84, 2, 1, 1, 2, 4, 1}, {65, 1, 1, 1, 1, 4, 2},
{12, 2, 2, 2, 2, 4, 2}, {57, 2, 1, 1, 2, 2, 2}, {34, 1, 1, 1, 1, 4, 2},
{42, 1, 2, 2, 2, 4, 2}, {57, 2, 1, 2, 1, 2, 2}, {28, 1, 1, 1, 1, 4, 2},
{67, 2, 2, 2, 2, 4, 1}, {54, 2, 1, 1, 1, 2, 1}, {51, 1, 1, 1, 2, 4, 2},
{35, 2, 1, 2, 1, 2, 2}, {29, 2, 1, 1, 1, 1, 2}, {31, 1, 1, 1, 1, 2, 1}

...

Ausblick

- Einbeziehung der Haushaltsstruktur (zuordnen von Haushalten)
- Zuordnen von Verkehrsverhalten und Mobilitätswerkzeugen zu jedem Agenten
- Dynamik integrieren in die soziodemographische Struktur der Agenten (altern, heiraten,...)
- entwickeln von besseren Methoden...

Referenzen

Smith L., Beckman R., Anson D., Nagel K. and Williams M. (1995) TRANSIMS: Transportation SIMulation System, *Los Alamos National Laboratory Unclassified Report*, LA-UR-95-1664, Los Alamos, NM 8744.

Axhausen, K. W. and Gärling T. (1991) Activity-based Approaches to Travel Analysis: Conceptual Frameworks, Models and Research Problems, *U.S. Department of Commerce, National Technical Information Service*, TSU Ref:628.

Gärling T., Kwan M. and Colledge R. G. (1994) Computational-process modeling of household activity scheduling, *Transportation Research*, B **28**, 355-364.

Papacostas C. S. and Prevedouros P. D. (1993) *Transportation Engineering and Planning*, (2ed Edition), Prentice Hall, Englewood Cliffs, NJ.

Anderson, E.B. (1997) *Introduction to the Statistical Analysis of Categorical Data*, Springer, Berlin.

Deming, W. E. and Stephan, F. F. (1940) On a least squares adjustment of a sampled frequency table when the expected marginal tables are known, *Annals of Mathematical Statistics* **11**, 427-444.

Referenzen

Voellmy, A., M. Vrtic, B. Raney, K.W. Axhausen, K. Nagel (2001) Status of a TRANSIMS implementation for Switzerland, *Networks and Spatial Economics*.

Beckman, R. J., K.A. Baggerly und M. D. McKay, (1996) Creating Synthetic Baseline Populations, *Transportation Research-A*, **30** (6), 415-429.

Ireland C. T. and Kullback S. (1968) Contingency tables with given marginals, *Biometrika* **55**, 179-188

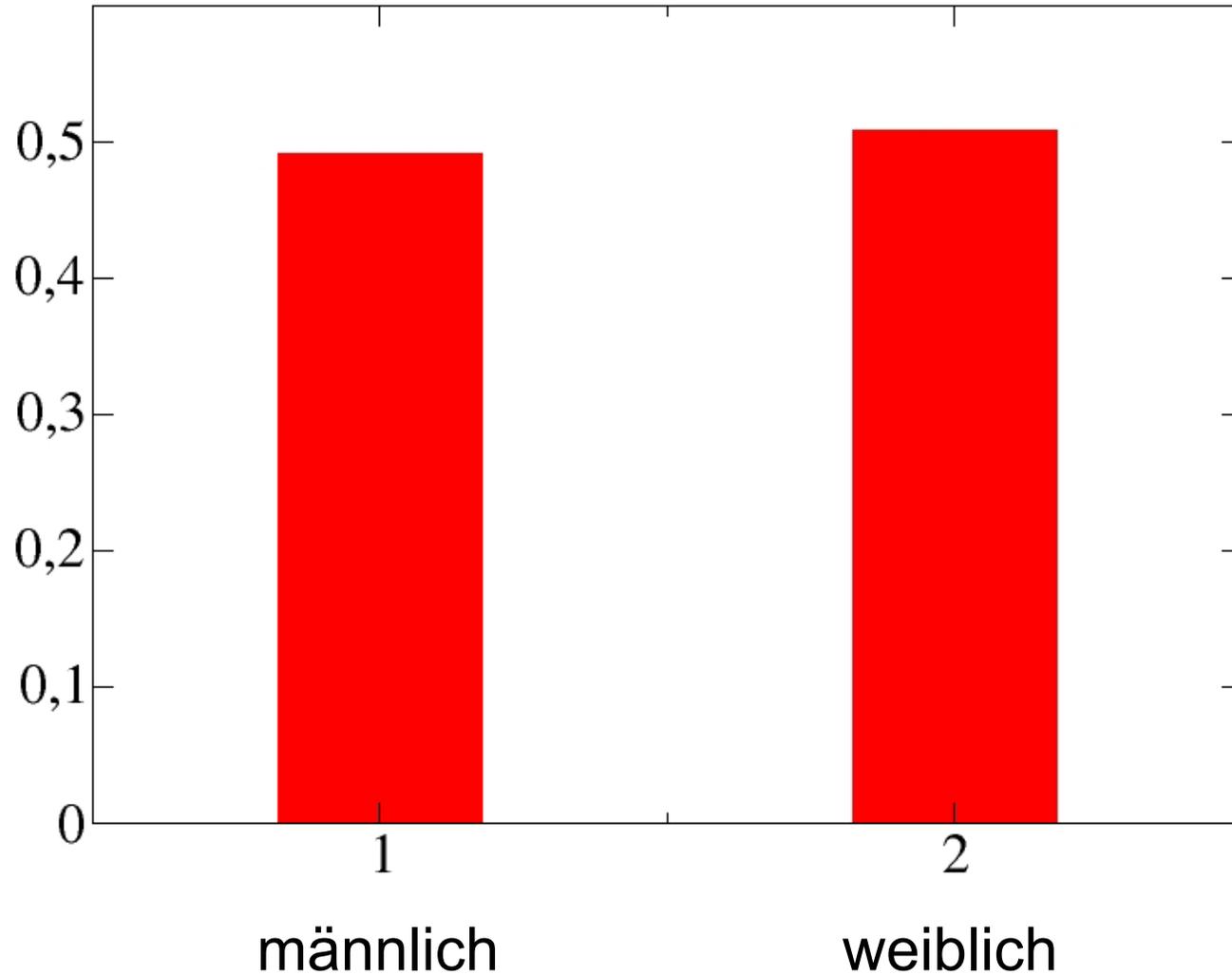
Bundesamt für Statistik (BFS) / Bundesamt für Raumentwicklung (ARE), (2001), Mikrozensus Verkehrsverhalten 2000, *Bundesamt für Statistik*, Bern, 2001.

Bundesamt für Statistik (BFS) / Bundesamt für Raumentwicklung (ARE), (1995), Mikrozensus Verkehrsverhalten 1994, *Bundesamt für Statistik*, Bern, 1995.

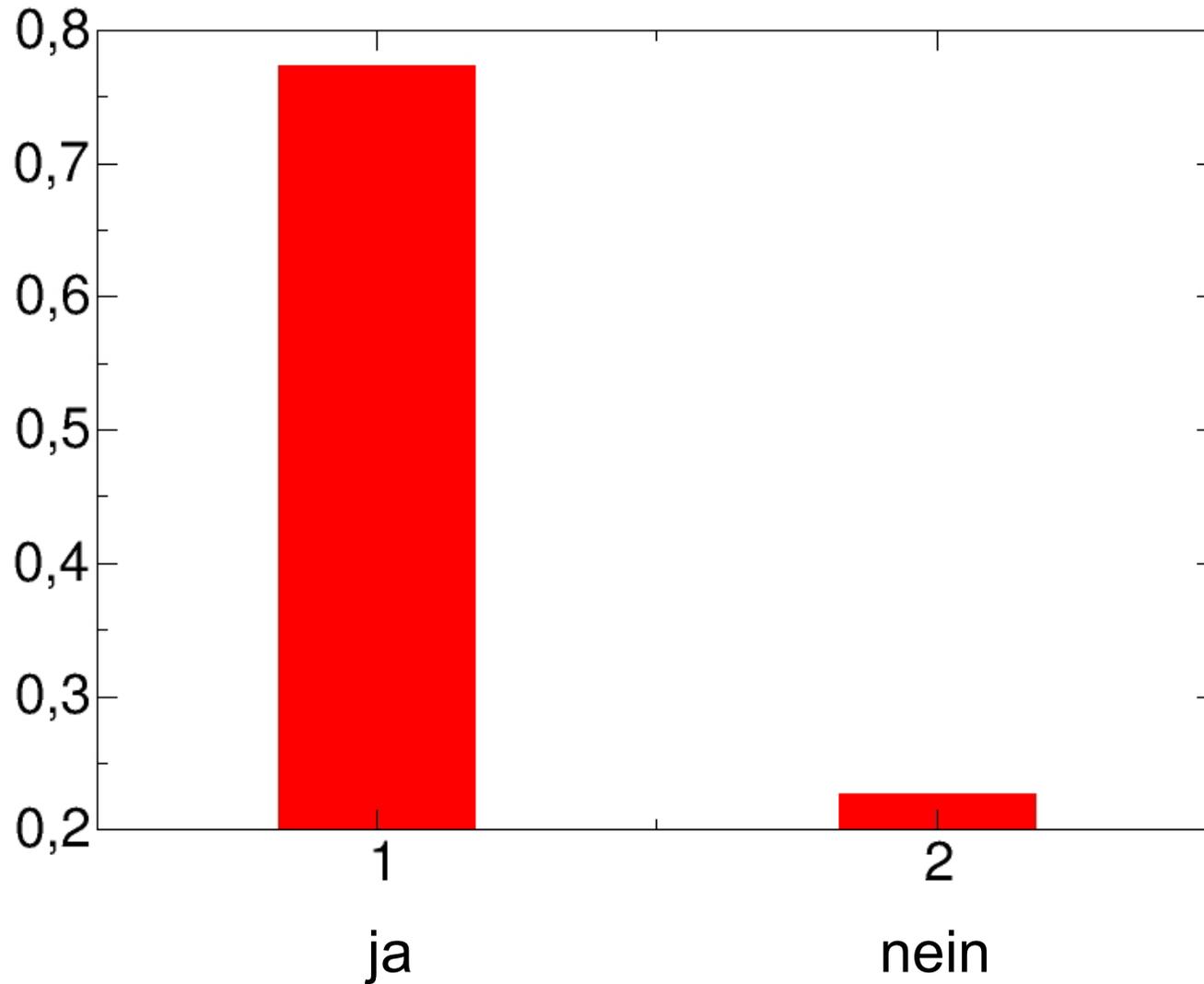
Bundesamt für Statistik (BFS) / Bundesamt für Raumentwicklung (ARE), (1990), Mikrozensus Verkehrsverhalten 1989, *Bundesamt für Statistik*, Bern, 1990.

Bundesamt für Statistik (BFS), (2001), Public Use Samples 1970, 1980, 1990, *Bundesamt für Statistik*, Bern, 2001.

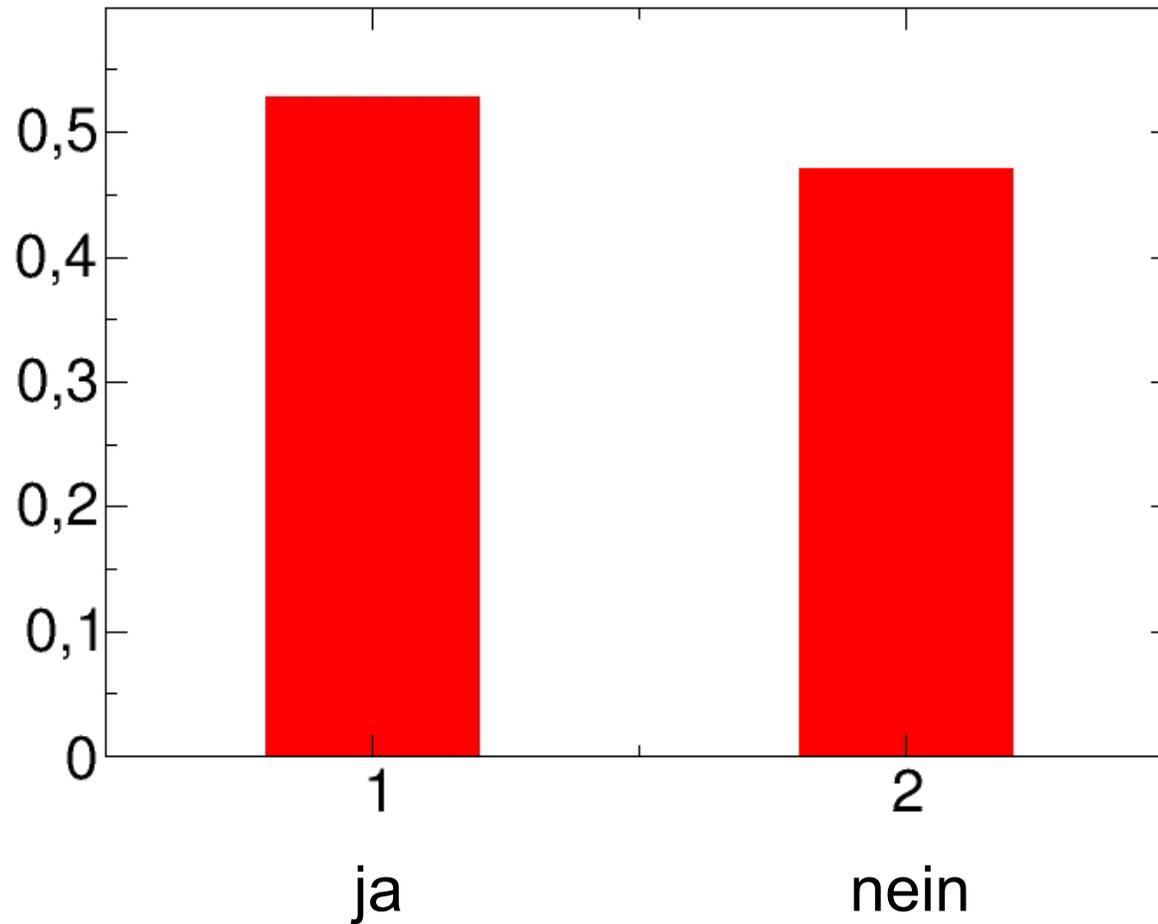
Anhang, Mikrozensus 2000 & VZ2000, Geschlecht



Anhang, MZ 2000 & VZ2000, Führerscheinbesitz



Anhang, Mikrozensus 2000 & VZ2000, Erwerbstätigkeit



Anhang, Mikrozensus 2000 & VZ2000, GA-Besitz

