

Adrian Herzog und Claudia Hofstetter

Strukturatlas des Kantons Zürich

Methodik

Zusammenfassung

Das vorliegende statistik.info beleuchtet die methodischen Aspekte des neuen Strukturatlases des Kantons Zürich. Wie jede derartige Kartensammlung soll auch dieser Atlas einen Erkenntnisgewinn über die Struktur seines Bezugsraums ermöglichen. Ein spezieller Fokus liegt hier beim Aufdecken und Darstellen von kleinräumigen Strukturen. Um dies zu ermöglichen wurde eine Methode auf einem neuartigen Konzept entwickelt, die auf den bereits veröffentlichten Daten der Umgebungsanalyse beruht und nun als Webapplikation vorliegt. Statt die Daten höchst detailliert und in ihrer Ausprägung sehr zufällig darzustellen (z.B. in einem ungeneralisierten Hektarraster) oder die Daten auf administrative Gebietseinheiten aufzuagggregieren und damit sehr viel an Detailinformation zu verlieren (z.B. als Gemeindekarten), betrachten wir für jeden Punkt eines feinen Rasters dessen Nachbarschaft: die konkrete Variablenausprägung wird mit Hilfe der Punkte aus seiner dynamisch definierten Nachbarschaft berechnet. Indem die Zahl der Nachbarn variiert wird, können bei einer kleinen Zahl von Nachbarn (z.B. 100 Personen) feine, im Falle von grossen Nachbarschaften (z.B. 5'000) gröbere Strukturen herausgearbeitet werden. Mit dieser Methode wird über das gesamte Gebiet eine einheitliche Informationsdichte und damit auch eine gleichmässige Zuverlässigkeit der Information erreicht. Etwas ungewohnt ist dabei, dass die dargestellten Werte auf unterschiedlich grossen Nachbarschaften beruhen (in Metern gemessen; sie betreffen aber immer die gleiche Zahl an Personen). Die Visualisierung der statistischen Oberflächen erfolgt als so genannte «Heatmaps» auf einem Hintergrund, der von Google Maps bereitgestellt wird. Für diesen Mashup mussten verschiedene kartographische und geometrische Probleme gelöst werden.

1. Einleitung

Der neu aufgebaute Strukturatlas des Kantons Zürich steht in einer Reihe von Webangeboten des Statistischen Amtes, die der Öffentlichkeit räumlich fein aufgelöste Informationen zur Verfügung stellen. Die demographische Umgebungsanalyse (2006) fand eine grosse Resonanz und wurde in den Medien rege diskutiert. In der Zwischenzeit wurde sie mit verschiedenen Datensätzen angereichert, die über die Demographie hinausgehen. Die Daten der Umgebungsanalyse bilden nun auch das Fundament für die vorliegende Webapplikation (<http://www.statistik.zh.ch/strukturatlas>). Während bei der Umgebungsanalyse die Betrachtungsrichtung des Anwenders auf einer spezifischen

Lokalität liegt, dessen Umgebung im Detail analysiert werden kann, versuchen wir mit dem «Strukturatlas» eine Übersicht über die räumliche Ausprägung der einzelnen Datensätze zu geben. Der Loc@lisator als weiteres Angebot unserer Website wiederum erlaubt es, aufgrund der persönlichen Präferenzen eines Benutzers die Teilgebiete des Kantons zu bewerten. Grundlage dafür sind gleiche und ähnliche Variablen, wie sie hier präsentiert werden.

Das vorliegende statistik.info versucht, diesen Strukturatlas in einen grösseren Zusammenhang zu stellen, das methodischen Vorgehen und die dabei auftauchenden Probleme im Schnittbereich Statistik und Kartographie zu beleuchten – speziell auch die verwendeten Generalisierungs- und Visualisierungskonzepte sowie die Integration in Google Maps.

2. Strukturatlanten

Ziel und Zweck eines Strukturatlases

Wie auch andere Strukturatlanten (vgl. z.B. Schuler et al. 1985, oder Schuler et al. 2007) beabsichtigt die vorliegende Kartensammlung, vorhandene räumliche Strukturen herauszuarbeiten. Sie ist in der gewählten räumlichen Auflösung eine erste umfangreiche Bestandesaufnahme zum Kanton Zürich. Ein spezielles Ziel liegt im Aufdecken von eher feinräumigen Strukturen. In der momentanen Ausgestaltung dieser Webapplikation liegt es beim Benutzer bzw. der Benutzerin, gleiche und unterschiedliche Strukturen bei verschiedenen Themen zu erkennen und diese einzuordnen versuchen. Erfahrungen mit unserer Kundschaft lassen vermuten, dass dies für viele am Kanton Zürich Interessierte reizvoll ist und auch ohne grössere Probleme machbar sein wird. In einem späteren Schritt werden wir versuchen, mit Methoden der räumlichen Statistik den Raum aufgrund der Ähnlichkeit der Variablenausprägungen zu gliedern.

Die räumlichen Unterschiede zeigen die Vielfältigkeit innerhalb des Kantons Zürich auf verschiedenen Ebenen: während bereits aufgrund von Gemeindekarten unterschiedliche Grossräume erkennbar sind, ist es nun mit diesem Atlas möglich, auch viel kleinräumigere Bereiche zu analysieren. Dabei zeigt sich gerade innerhalb der städtischen Gebiete, dass mit diesem Instrument auch in sehr lokale Bereiche vorgestossen werden kann.

Die Webapplikation «Strukturatlas» ist somit ein Informations- und Arbeitsinstrument, das es zunächst dem Benutzenden überlässt, Muster zu erkennen, Segregationstendenzen zu suchen und verschiedene Teilräume miteinander zu vergleichen. Sie lässt Raum, um Unterschiede zu interpretieren und zu bewerten. Dadurch dass die Daten so präsentiert werden, dass über den ganzen Raum hinweg eine einheitliche Informationsdichte garantiert ist, sind Fehlschlüsse, wie sie häufig in anderem Kontext durch Zufallseffekte oder ungünstig visualisierte Daten möglich sind, weitgehend ausgeschlossen. Dies gilt nicht zuletzt auch für die Betrachtung grossräumiger Raumstrukturen, bei denen im Falle der Gemeindekarten häufig durch relativ zufällige Ausprägungen in Kleingemeinden der visuelle Gesamteindruck stark gestört werden kann. Die einheitliche Generalisierung des Strukturatlases verhindert Fehlinterpretationen in dieser Hinsicht weitgehend.

Im Folgenden soll gezeigt werden, welche methodischen Überlegungen hinter dem Vorhaben standen und wie wir diese Webapplikation konkret umgesetzt haben.

3. Kartographisches

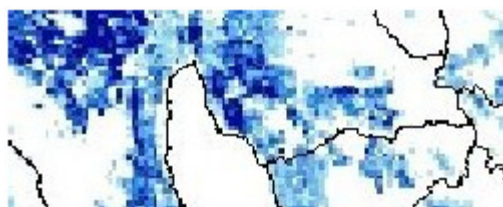
Ausgangsdaten: ungeneralisierte Information auf Hektarbasis

Bei den für den Strukturatlas verwendeten Rohdaten handelt es sich in den meisten Fällen um Punkthinformationen (Erhebungen wie Volks- und Betriebszählungen und Registerauswertungen wie das virtuelle Einwohnerregister VESTA), welche zunächst hektarweise aggregiert wurden. In der Regel beschreiben die Karten Relativvariablen, die durch die Verknüpfung verschiedener Datensätze errechnet wurden, wie z.B. den Anteil einer bestimmten Altersklasse an der Gesamtbevölkerung. Da die Grössen der Relativwerte je Rasterzelle auch als Höhen interpretiert werden können, spricht man in diesem Zusammenhang häufig auch von statistischen Oberflächen. Die Visualisierung geschieht entweder als eingefärbtes Raster oder seltener auch als eigentliche 3D-Darstellung der Oberfläche.

Hier eine einfache Darstellung einer für den Strukturatlas typischen Rohvariablen:

Grafik 1: Daten auf Hektarbasis

Bevölkerungsverteilung 2000



Die einzelnen Zellen des über den Kanton Zürich gelegten 100-Meter-Rasters repräsentieren sehr unterschiedlich viele Punkte (bzw. Personen, Betriebe, Beschäftigte etc.). Damit erhält die Oberfläche ein sehr raues, unruhiges Aussehen. Die einzelnen Werte sind dabei statistisch betrachtet sehr unterschiedlich zuverlässig: Im Extremfall zeigt eine Hektare den Wert für eine einzelne Person, womit dort z.B. der Anteil der Altersklasse 0 oder 100% beträgt. Diese Heterogenität und Zufälligkeit der Information – allgemein das Problem der kleinen Zahlen – führt dazu, dass es für einen Benutzer unmöglich ist, die Zuverlässigkeit der dargestellten Ausprägungen zu beurteilen. Dieses «Datenrauschen» erschwert oder verunmöglicht es, Strukturen zu erkennen und macht eine Generalisierung unabdingbar.¹

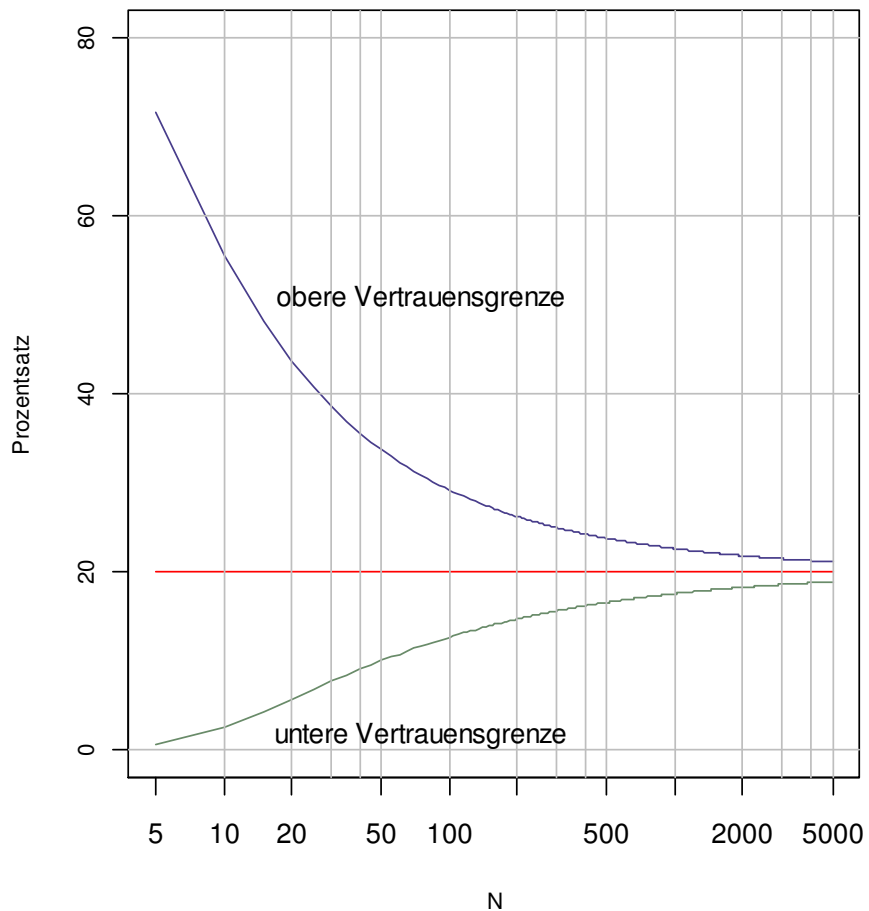
Die im Strukturatlas am häufigsten visualisierte Art sind Anteilswerte (z.B. der Anteil einer Bevölkerungsgruppe an der Gesamtbevölkerung). Die Abschätzung der Unsicherheit einer Proportion kann statistisch einfach mit der die Breite ihres Vertrauensbereichs² erfolgen. Bei einer angenommenen Proportion von 20% ergeben sich folgende untere und obere Konfidenzintervalle bei einer unterschiedlich grossen Zahl von N (der Zahl Personen, auf die sich der Prozentsatz bezieht):

¹ Bei stetiger und zuverlässiger Information stellt sich das Generalisierungsproblem kaum, derartige Oberflächen (wie z.B. die räumliche Verteilung von Temperatur, Schadstoffen, Sonneneinstrahlung etc.) können problemlos auch mit einer maximalen Auflösung visualisiert werden. Zu Vergleichszwecken finden sich auch einige derartige Oberflächen im Strukturatlas.

² Wir verwenden hier den exakten Test von Clopper und Pearson.

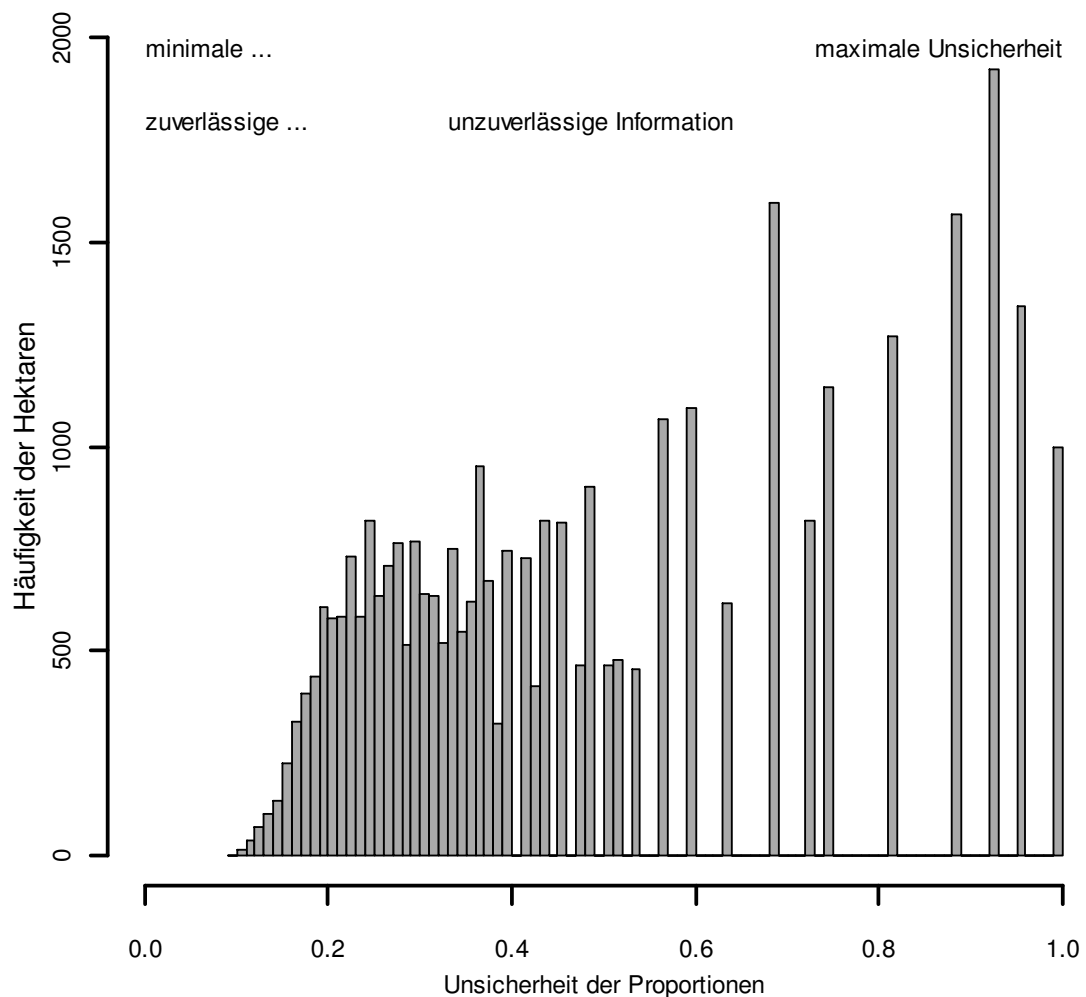
Grafik 2: Konfidenzintervalle von Prozentsätzen

Beispiel mit $p = 20\%$



Umgelegt auf hektarweise Bevölkerungsinformation für den Kanton Zürich ergibt sich die Verteilung der Informationsunsicherheit, wie in Grafik 3. Die aller kleinste Zahl von Hektaren zeigt einen zuverlässigen Prozentwert (repräsentiert durch die relativ kurzen Balken ganz links aussen), viele Angaben sind komplett unsicher (in einem Fünftel der bevölkerten Hektaren kann ein ausgewiesener Prozentsatz von 20% irgendwo im Bereich von 0-80% liegen).

Grafik 3: Informationsunsicherheit im Hektarraster



Kartographische Generalisierung

Es ist somit unumgänglich, die statistische Oberfläche in ihrer Rohform in irgendeiner Form zu generalisieren. Damit soll eine höhere Zuverlässigkeit der Information über das Gebiet hinweg erreicht werden, die es dann erlaubt, die zugrunde liegende räumliche Verteilung sauber zu kommunizieren. So wird es für den Benutzer der Karte erst möglich, die visualisierte Oberfläche mit einer gewissen Sicherheit zu beurteilen.

Damit stehen zwei Anliegen in einer starken Konkurrenz: einerseits möchte man die Informationen möglichst detailliert darstellen (z.B. auf Niveau Hektaren), andererseits sollte die dargestellte Information möglichst zuverlässig sein. Während im ersten Fall die einzelnen unzuverlässigen Informationen das Gesamtbild derart stören und damit das Erkennen von Strukturen verunmöglichen (man sieht vor lauter Bäumen den Wald nicht), verliert man im anderen Fall derart viel an spezifischen lokalen Informationen, dass die kommunizierte Informationsmenge minimal wird.

Wir haben es damit mit einem Problem der kartographischen Generalisierung zu tun, bei der es um «Vereinfachen im Rahmen einer inhaltlich gegebenen Konzeption» (Imhof 1986,66) geht. Dieses Vereinfachen kann mit folgenden elementaren Vorgängen

beschrieben werden: «Vereinfachen – Vergrössern - Verdrängen – Zusammenfassen (Aggregieren) – Auswählen (Selektieren) – Klassifizieren - Bewerten» (Hake 2002, 168). Zentrales Verfahren wird hier eine Form der Aggregation sein: die zugrundeliegenden Daten werden soweit aufbereitet bzw. aufaggregiert, dass deren kartographische Umsetzung die Anforderung für eine zuverlässige Kommunikation erfüllt.

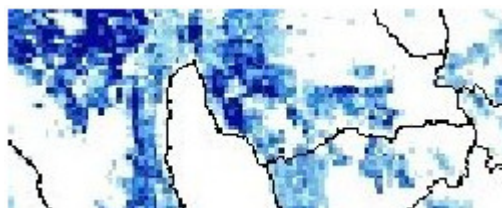
Aggregation in gröbere Raster

Verbreitetes Verfahren für die Aggregation ist die Vergröberung der räumlichen Auflösung der Bezugsgebiete: statt Hektaren werden grössere Rasterzellen verwendet. Grafik 1 Grafik 4 zeigt die generalisierende Wirkung der Aggregation:

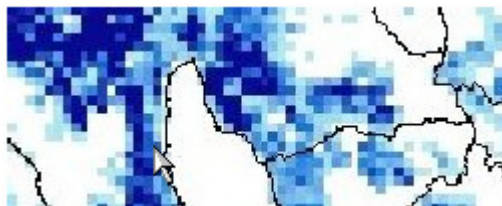
Grafik 4: Generalisierung durch Aggregation

Aggregation von 1.2 Mio. Punkten auf unterschiedlich feine Gebietseinheiten (Bevölkerungsverteilung 2000 (Ausschnitt))

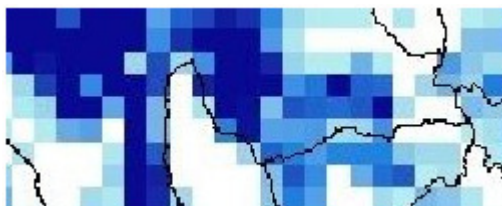
100x100m Rasterzellen, Hektarraster



200x200m Rasterzellen

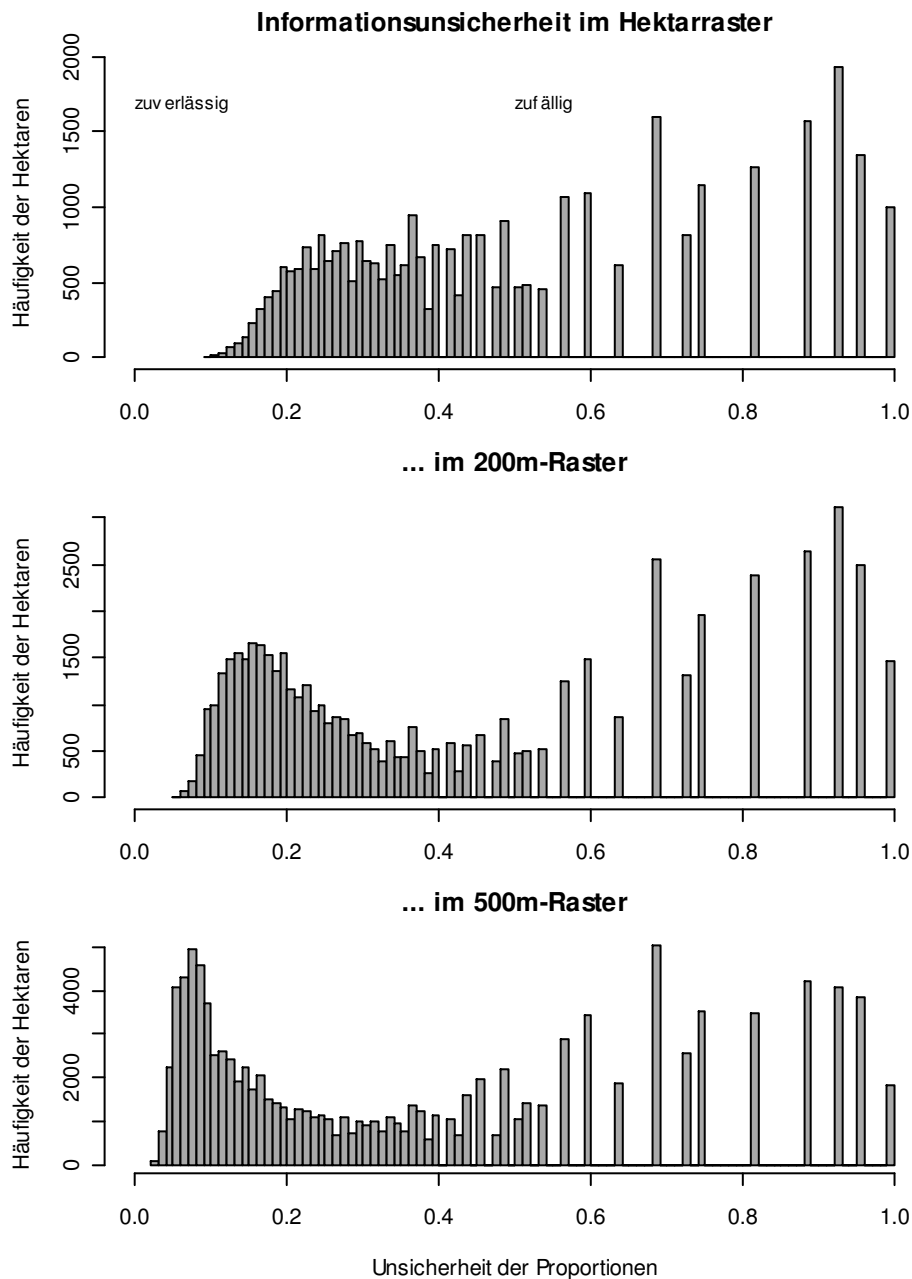


500x500m Rasterzellen



Durch das Verändern der Rasterweite wird die Information in vielen Teilgebieten tendenziell zuverlässiger, andererseits bleibt die grundsätzlich beschriebene Problematik erhalten, indem immer noch Rasterzellen mit extrem wenigen Beobachtungen auftreten können und dadurch weiterhin in grossen Gebieten der dargestellte Prozentsatz sehr zufällig ausfallen kann.

Grafik 5: Informationsunsicherheit in aggregierten Rastern



Zu beachten ist, dass die optimale Auflösung stark vom Inhalt der Oberfläche abhängt: Variablen aus der Bevölkerungsstatistik könnten dank der viel grösseren Fallzahlen viel detaillierter aufbereitet werden als beispielsweise Variablen, die aus der Betriebszählung mit ihren kleineren Grundzahlen abgeleitet werden.

Aggregation in administrativen Einheiten

Neben der rasterweisen Aggregation ist die Aggregation in unregelmässige Umrisse eine weitere Variante: entweder in administrativ begründete Einteilungen (z.B. Gemeinden, Bezirke, Regionen) oder in ad hoc gebildete Umrisse. Beim letzteren Verfahren wird versucht, Gebiete möglichst ähnlicher Ausprägung auszuscheiden und damit eine so genannte dasymetrische Karte zu erreichen.

Bei der Aggregation zu jeder wie auch immer gearteten Gebietsaufteilung gilt es ein fundamentales Problem zu beachten: das von Openshaw (1984) als «**modifiable areal unit problem**» (MAUP) beschriebene Phänomen, dass die Resultate statistischer Analysen unter Umständen stark von der für die Aggregation verwendeten Raumgliederung abhängen können. Dies gilt im gleichen Masse auch für das Resultat einer kartographischen Visualisierung, indem die sichtbaren Raumstrukturen davon abhängen, wie die einzelnen Elemente (Personen, Gebäude, Beschäftigte etc.) zu Gebieten aggregiert werden bzw. wo genau die Gebietsgrenzen gezogen werden. Dadurch werden statistische Werte und auch die daraus abgeleiteten Karten in einem gewissen Grad zufällig.

Da es sich beim Kanton Zürich um ein relativ kleines Gebiet handelt, sind bereits die 171 Gemeinden eine zu starke Aggregation für die Visualisierung von Raumstrukturen, was die Gemeindekarte mit der gleichen Information wie in den obigen Beispielen zeigt:

Grafik 6: Generalisierung durch Aggregation in administrative Einheiten

Ausschnitt aus einer Gemeindekarte (Bevölkerungsdichte 2000)



Die Verbesserung der Zuverlässigkeit der Information geht hier auf Kosten des räumlichen Details und wird den Anforderungen, wie wir sie hier für die Möglichkeit kleinräumiger Aussagen stellen, nicht gerecht. Auch bei solchen – sehr verbreiteten – Choroplethenkarten ist zudem eine beträchtliche Variabilität der Zuverlässigkeit anzutreffen (bei der kleinsten Gemeinde können je nach Thema ebenfalls nur sehr kleine Werte vorkommen; z.B. hat es bei der Betriebszählung 2005 in knapp 20 Gemeinden des Kantons weniger als 25 Arbeitsstätten).

Unsichere Information

Eine quantitative Abschätzung der Unsicherheit der Information³ in den bisher gezeigten Generalisierungsbeispielen (und gleich auch die entsprechenden Kurven für die hier entwickelten «Umgebungen») zeigt die Grafik 7. Bei den Rastern zeigt ein beträchtlicher Teil des dargestellten Gebiets ziemlich unsichere Werte. Die Schnittpunkte beispielsweise mit der statistischen Oberfläche, bei der 100 Personen hinter einem berechneten Wert stehen, liegen weit links, d.h. nur wenige Zellen der regelmässigen Raster weisen mehr als 100 Personen auf. Damit wird auf der weitaus grössten Fläche der Karten jeweils keine oder nur sehr unzuverlässige Information dargestellt⁴.

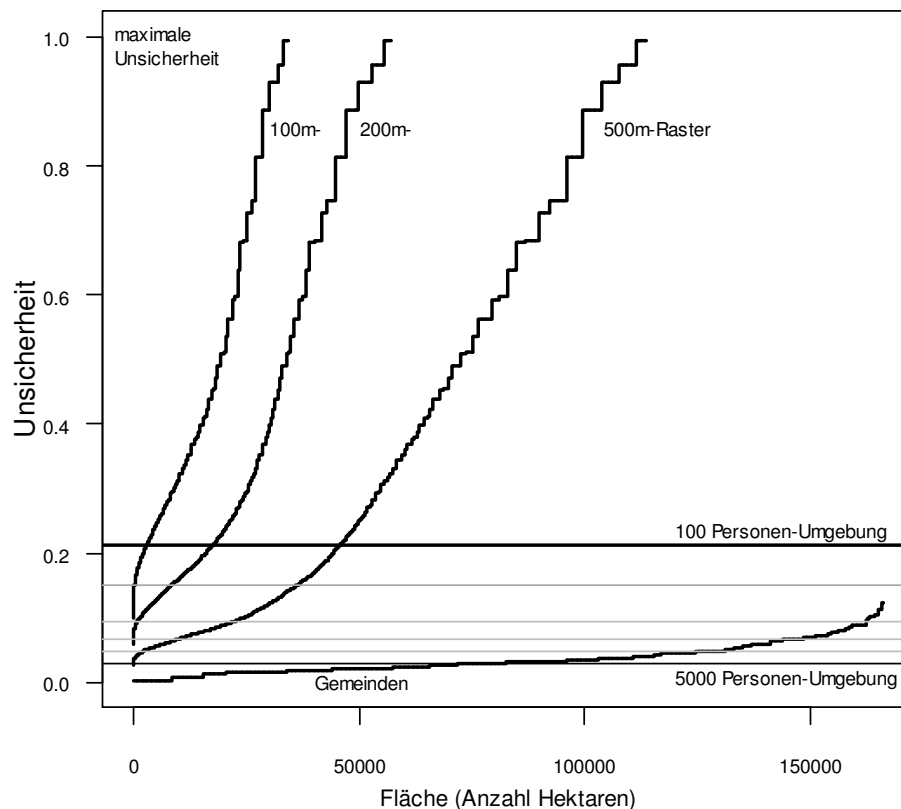
³ Gerechnet wurde mit einer Proportion über das ganze Gebiet von wiederum 20%. Damit kann für jedes Teilgebiet die Breite des Vertrauensbereichs berechnet werden.

⁴ Die Nachbarschaftsumgebungen mit ihrer fixen Grösse der Zahl der Nachbarn finden sich als horizontale Linien: sie haben über das ganze Gebiet eine einheitliche Zuverlässigkeit, die Umgebungsraster mit weniger Personen eine kleinere, die stärker generalisierten eine grössere. Zudem decken sie das ganze Gebiet des Kantons ab.

Die Gemeindekarte zeigt in diesem Zusammenhang eine sehr grosse Zuverlässigkeit – allerdings natürlich auf Kosten einer extrem geringen räumlichen Auflösung bzw. Informationsmenge.

Grafik 7: Unsicherheit in unterschiedlich generalisierten Karten

100, 200, 500m-Raster und Gemeindekarte



Das im Strukturatlas verwendete Verfahren: Nachbarschaften

Für den Strukturatlas haben wir eine Lösung gesucht, die oben erwähnten Probleme der Gebietsaufteilung umgeht und dem eigentlichen Inhalt der Daten und den Aussagen besser gerecht werden kann. Gesucht war ein neuer Kompromiss zwischen möglichst detaillierter und trotzdem zuverlässig generalisierter Information.

Wichtigstes Ziel beim Strukturatlas war es, eine einheitliche und gleichmässige Informationsdichte über das ganze Gebiet zu erreichen. Die Daten aus der Umgebungsanalyse und vor allem die dort entwickelten Konzepte eignen sich für diesen Anspruch ausgezeichnet: Die in der Umgebungsanalyse ausgewiesenen Werte beziehen sich nicht auf eine irgendwie geometrisch definierte Nachbarschaft, vielmehr steht die uniforme Informationsdichte im Vordergrund: jeder ausgewiesene Wert wird individuell

aufgrund einer vorgegebenen Zahl von Nachbarschaftspunkten berechnet. Der Umkreis, innerhalb dessen diese Zahl von Punkten angetroffen wird, ist zwangsläufig variabel.

Das Nachbarschaftskonzept hat den Vorteil, dass es sehr einfach verständlich ist: der Indikator bezieht sich auf die nächsten k Nachbarn eines Punktes. Dieselbe Datenstruktur ist auch die Basis der Umgebungsanalyse⁵:

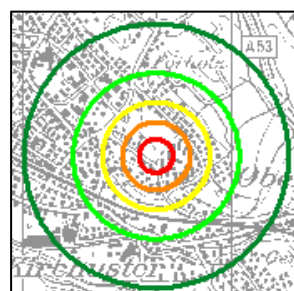
Grafik 8: Die k nächsten Nachbarn in der Umgebungsanalyse

3 Nachbarschaftstypen: Personen, Gebäude und Beschäftigte

Radius der Umgebung	466 m	638 m	828 m	1372 m	1938 m
Personen im Jahr 2008	200	500	1000	2000	5000

Radius der Umgebung	359 m	545 m	852 m	1300 m	3164 m
Anzahl Gebäude im Jahr 2000	200	500	1000	2000	5000

Radius der Umgebung	188 m	260 m	344 m	549 m	978 m
Anzahl Beschäftigte im Jahr 2005	200	500	1000	2000	5000



<http://www.statistik.zh.ch/umgebungsanalyse>

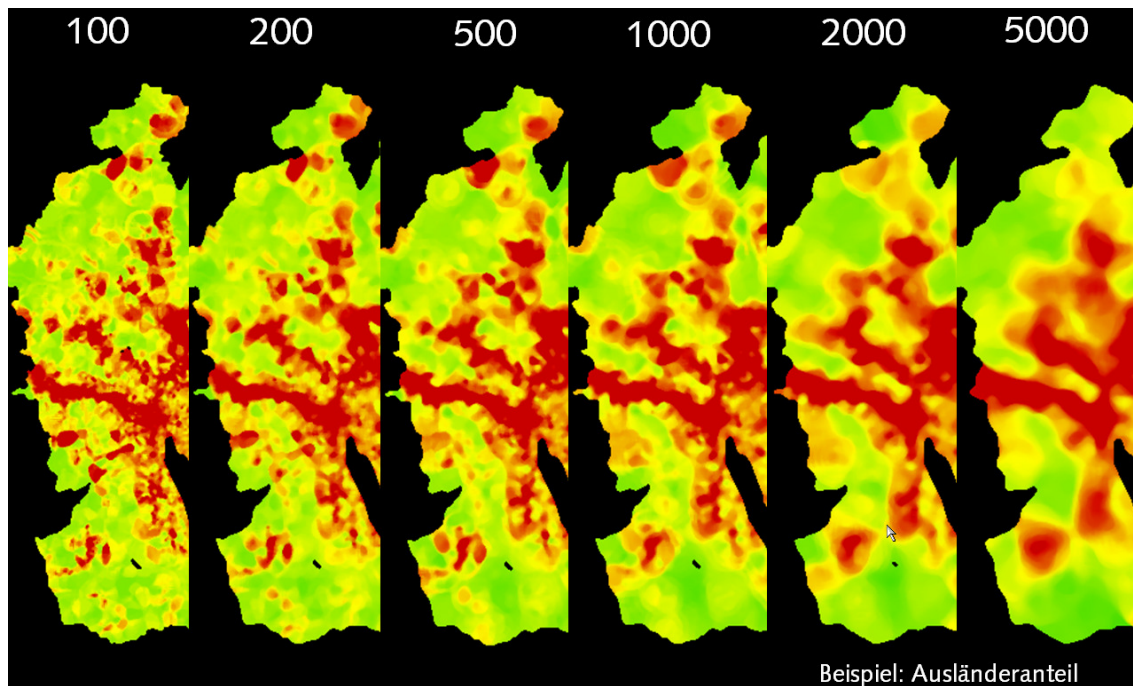
Derartige Verfahren, welche für eine Analyse die nächsten k Nachbarnpunkte zusammensuchen, sind in der räumlichen Datenanalyse nicht unüblich. Die Erarbeitung von Nachbarschaftsstrukturen ist häufig die Voraussetzung für die weitere statistische Analyse von Punktinformationen (räumliche Autokorrelation, Klassifikationen etc.; vgl. Bivand 2008, 239ff.; im GIS-Umfeld spricht man von fokaler Nachbarschaftsstatik). Derartige k-nächste-Nachbarn-Algorithmen werden in der Regel aber selten für derart grosse Datenmengen ($>10^6$) und Aggregationsniveaus ($100 \leq k \leq 5'000$) angewendet. Die Berechnung der Variablen wurde erst durch ein entsprechendes Vorverarbeitung der Rohdaten (Aggregation in Hektarrastern) und einer anschliessenden Interpolation möglich (vgl. Herzog und Hofstetter, 2009). Der implizit resultierende Genauigkeitsverlust durch das Interpolieren kann in seinen Auswirkungen vernachlässigt werden.

Die Berechnung der k nächsten Nachbarn erfolgt im Rahmen des Strukturatlases ungewichtet. Mit der Festlegung verschiedener k-Werte (100, 200, 500, 1000, 2000, 5000) wird eine unterschiedlich starke Generalisierung erreicht, wie dies Grafik 9 zeigt:

⁵ Die Methodik für die Aufbereitung der Daten besteht in einem aufwändigen Preprocessing mit Hilfe des Geographischen Informationssystems ArcGIS, bei dem zunächst die Werte für geometrische Nachbarschaften extrahiert wurden. Anschliessend werden Interpolationen vorgenommen, im Falle der Umgebungsanalyse für einzelne Punkte und online mit Visual Basic-Programmen. Hier beim Strukturatlas mussten alle Punkte des Kantons berechnet werden, wobei lokal ausgeführte Java-Programme zum Einsatz kamen. Die algorithmischen Details finden sich bei Herzog und Hofstetter (2009, 9ff.).

Grafik 9: Generalisierungswirkung als Funktion der Umgebungsweite

Beispiel: Ausländeranteil



Die unterschiedlichen Generalisierungsstufen verdeutlichen Raumstrukturen verschiedener Massstabsebenen, von lokalen Besonderheiten bei kleinem k bis zu grossräumigen Strukturen im stark generalisierten Fall.

Neben der verwendeten Methode der k nächsten Nachbarn wären auch noch andere Verfahren möglich gewesen, deren Verhalten aber mit der hier zu bearbeiteten Zahl von Beobachtungen nicht abgeschätzt werden kann. Es wäre dies z.B. die Hot Spot Analysis (Getis-Ord Gi*) oder das Verfahren der Kerndichteschätzung (*kernel density estimation*), welche für die reinen Punktdichtekarten zur Erstellung von Dichtekarten angewendet werden könnte. Diese Verfahren werden als dynamisch generierte Mashups verschiedentlich eingesetzt⁶ und werden häufig als *heatmaps* bezeichnet. Die übliche Zahl der dabei verarbeiteten Punkte ist aber vergleichsweise klein. Für die Spezialfälle mit den Auswertungen für die nächsten zentralörtlichen Einrichtungen etc. wären von einem ästhetischen Standpunkt aus bei Distanzkarten solche kernel-density-Lösungen schöne. Inhaltlich ist die Lösung der k nächsten Einrichtungen aber korrekt, sie entspricht konzeptionell den übrigen Karten und ist viel einfacher kommunizierbar.

Die immer gleiche grosse Zahl von Beobachtungen (Personen, Gebäude etc.), die hinter jedem Punkt der Karte stehen, wird aufgrund von sehr unterschiedlich grossen Suchradien erreicht und berechnet. In dünn besetzten Gebieten errechnet sich der dargestellte Wert aus einer sehr grossen Umgebung, was dazu führt, dass relativ weit entfernte Siedlungen ebenfalls Einfluss haben. Das würde für ein Ausblenden solcher informationsarmen Gebiete sprechen (vgl. unten die Ausführungen über mögliche Varianten der Gebietsabdeckung).

Die verwendete Methodik eignet sich speziell auch aus Datenschutzgründen, weil es nicht möglich ist, auf einzelne Punkte bzw. Personen rückzuschliessen. Deswegen wer-

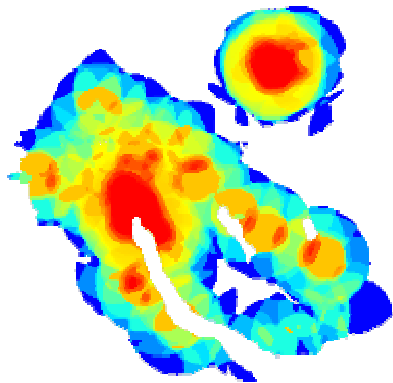
⁶ z.B. <http://www.heatmapapi.com>

den ähnliche Methoden z.B. auch für die Visualisierung von Kriminalität verwendet (vgl. BBC 2009).

Dadurch dass sehr grosse k-Werte verwendet werden, werden quasi kontinuierliche Übergänge erreicht. Dies gilt nicht für Nachbarschaften, die sich auf sehr wenige Werte (S-Bahnhaltestellen, Warenhäuser etc.) beziehen (vgl. Grafik 11). Dort ist die Oberfläche naturgemäss nicht stetig.

Grafik 10: Suche nach wenigen Punkten

Nähe zu wenigen zentralörtlichen Einrichtungen (3 Museen)

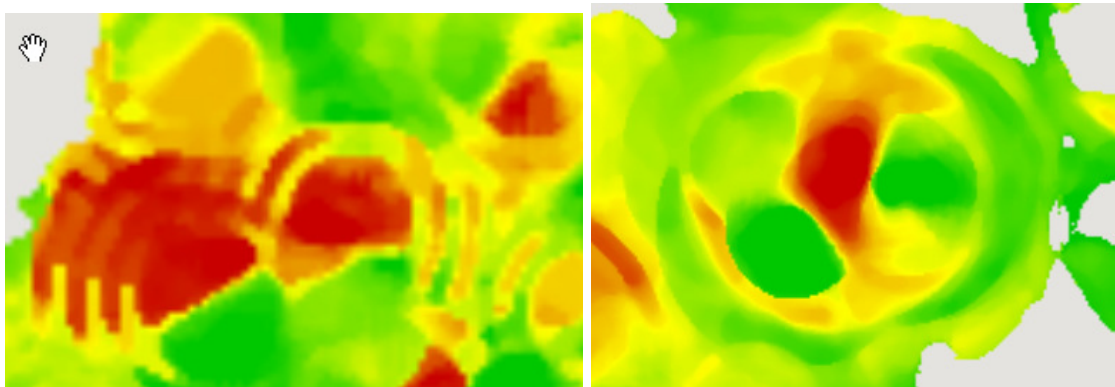


Spezielles zum verwendeten Vorgehen

Eigentliche Artefakte haben sich ergeben, weil wir beim Preprocessing nicht kontinuierlich Daten für primäre Umgebungen aufbereitet haben. Wir haben uns aus Aufwandgründen darauf beschränken müssen, für Hektarumgebungen mit folgenden Zellenweiten Basisdaten zu erstellen: in Einzelschritten von 1 bis 10, dann für 15, 20, 25 und 50 Umgebungszellen. Dadurch dass bei sehr geringen Datendichten Werte im Bereich von 1500-5000 Metern auf einer relativ kleinen Basis interpoliert werden müssen, ergeben sich in gewissen Fällen Artefakte (vgl. Grafik 11), die das Bild etwas stören. Wir vermuten auch, dass dieser Effekt in Situationen verstärkt wird, bei dem die Punkte an einzelnen Standorten stark konzentriert sind (z.B. bei den Beschäftigten).

Grafik 11: Artefakte

bei der Suche nach sehr wenigen Standorten bzw. in Gebieten mit geringer Informationsdichte



Die Ausprägung der Nachbarschaften leiden bei einigen Themen (noch) darunter, dass bei verschiedenen Variablen ausserhalb des Kantons keine Daten vorhanden sind. Dadurch ergibt sich (z.B. bei den Bevölkerungsdichtekarten) ein falscher Eindruck an der Kantonsgrenze, wo dichter besiedelte Gebieten in den Nachbarkantonen liegen. Gut beobachtbar ist das beispielsweise an der Grenze zu den Kantonen Schaffhausen und Zug. Bei den neueren Variablen (z.B. aus der aktuellen Betriebszählung) wurden deshalb auch die Werte der Nachbarkantone in die Basisberechnungen miteinbezogen.

4. Visualisierung, kartographische Umsetzung

Symbolisierung

Im Strukturatlas haben wir unterschiedliche Daten- bzw. Kartentypen verwendet, es sind dies Anteilswerte, Distanzen und schliesslich Veränderungsraten und andere Indizes.

Um die Unterschiedlichkeit dieser Konzepte zu verdeutlichen, haben wir versucht, sie je mit einem anderen Farbkonzept darzustellen. Wie die meisten der publizierten derartigen Karten verwenden wir Farbschemata, bei denen von einem heiss-kalt-Schema ausgegangen wird. Für «Heisse» Gebiete mit hohem Anteil, hoher Dichte etc. werden Rottöne verwendet, am anderen, kalten Ende sind es Blau- oder Grüntöne⁷. Ausgegangen sind wir von Farbkeilen aus ArcMap.⁸

Es zeigt sich, dass gut perzeptierbare und visuell ansprechende Farbübergänge, die zudem auch noch mit dem eher grünlichen Hintergrund (Vegetation im Satellitenbild) harmonieren, nicht einfach zu finden sind. Der häufig verwendete Übergang von grün über gelb nach rot eignet sich recht gut, hat aber den schwerwiegenden Nachteil, dass er für Personen mit Rot-Grün-Sehschwäche absolut ungeeignet ist.

Lösungen mit einem nur einem Farbübergang (z.B. von Gelb zu Braun) haben visuell nicht überzeugt, so dass wir schliesslich Übergänge mit einer weiteren Stützpunktfarbe (meist Gelb) verwendet haben.

Klassifikation

Um möglichst wenig an Information zu verlieren, haben wir die Datenwerte nicht klassiert. Da aber eine reine Umsetzung des Datenbereichs von Minimum zum Maximum stark von möglicherweise etwas zufälligen Extremwerten abhängen würde, haben wir alle Werte, die unterhalb des 5%-Quantils bzw. die oberhalb des 5%-Quantils liegen, jeweils in eine Klasse der Minimum- bzw. Maximumwerte gelegt. Die übrigen 90% der Werte, die innerhalb dieser Grenzen liegen, werden mit Hilfe eines kontinuierlichen Farbverlaufs umgesetzt. Dieses Vorgehen legt den Schwerpunkt auf der Optimierung der einzelnen Karte, was den Vergleich zwischen den Themen und den Generalisierungsstufen allenfalls etwas erschweren kann.

Die **Anteilswerte** zeigen in der Regel Prozentwerte einer Teilgruppe an der Gesamtzahl von Personen, Beschäftigten etc. Beispiele sind etwa: Anteil der über 80-Jährigen an der Bevölkerung, Haushalte mit 3 Personen, Wohngebäude gebaut 1981-90, etc.):

⁷ Dies im Gegensatz zur Localisator-Anwendung, wo vom Gegensatz gut-schlecht bzw. geeignet vs. nicht geeignet ausgegangen wird und eher eine Verkehrsampellogik verwendet wird (grün: go, rot: stop).

⁸ Die Interpolation von Farben zwischen zwei als Stützpunkten funktionierenden Farben erfolgt über eine Interpolation im Farbraum. Beachtet werden muss dabei, dass je nach verwendetem Farbraum (z.B. RGB [rot/grün/blau] oder HLS [*hue/lightness/saturation*, Farbton/Helligkeit/Sättigung] ziemlich unterschiedliche Keile erzielt werden.

Anteilswerte:
(wenig – viel)



Die **Distanzkarten** zeigen, welchen Radius der Kreis hat, bei dem man genau k Personen etc. antrifft. Als Variante präsentieren wir auch Karten, bei denen statt dem Radius des Kreises mit der Fläche des Kreises innerhalb einer k-Personenumgebung eine Dichte berechnet wird: $k / (\text{Fläche des Kreises})$. Da Radius und Fläche eines Kreises in einem quadratischen Zusammenhang stehen, verändert sich dadurch lediglich die Geschwindigkeit des Übergangs im Farbkeil.

Distanzen, Dichten:
(eng/dicht – weit/dünn)



Weitere Indizes wurden auf verschiedene Arten gerechnet und werden farblich auch unterschiedlich umgesetzt.

Veränderungen:
(Abnahme – Zunahme)



Walkability und Cyclability:
(gut - schlecht)



Schadstoffe:
(stark – schwach belastet)



Sonneneinstrahlung:
(hoch – tief)



Weitere Relativvariablen: Wohnfläche, Mietpreise

In manchen Fällen wäre es allenfalls noch wünschenswert, wenn die Farbschemata interaktiv zugewiesen oder verändert werden könnten. Dies wäre nur möglich, wenn weitere Berechnungsergebnisse datenbankmässig gespeichert würden und dann erst diese direkt in farbige Karten umgewandelt würden – quasi durch eine Substitution von Rechenleistung durch Speicherplatz, was durchaus möglich wäre, wenngleich auch gewisse verzögerte Antwortzeiten wahrscheinlich wären. Da gilt es die Erfahrungen mit dem Instrument bzw. entsprechendes Feedback abzuwarten.

Die Verwendung von Google Maps

Zur Visualisierung unserer statistischen Oberflächen haben wir uns entschieden, einen *Mashup* mit Google Maps zu machen, d.h. wir verwenden die Google Map-API (2009) und überlagern die von Google gelieferten Karten oder Satellitenbilder mit unseren Strukturdaten.

Die Kombination mit Satellitenbildern anstelle reiner thematischer Karten vereinfacht die Orientierung und lässt die statistische Oberfläche weniger abstrakt erscheinen. Auf zusätzliche Beschriftungen kann verzichtet werden, da sich der Benutzer mit Hilfe der Basisfunktionalität von Google Maps (Kartenhintergrund oder hybride Darstellung von Satellitenbild und Beschriftungen) problemlos orientieren kann. Etwas störend sind die farbigen Beschriftungen der Autobahnen (bei der Hybridvariante Satellitenbild/Karte),

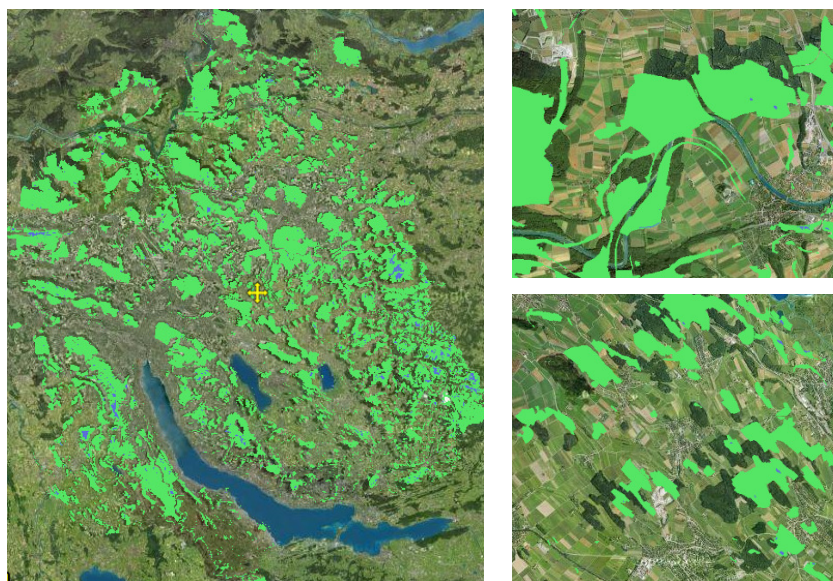
die die Aufmerksamkeit unnötigerweise auf sich lenken. Als dritte hier angebotene Möglichkeit kann auch auf den Kartenhintergrund komplett verzichtet werden.

Überlagerungsprobleme

Spezielle Probleme ergaben sich bei der Überlagerung der Oberflächenraster auf die Google-Karte. Das schweizerische Landeskoordinatensystem basiert auf einer schiefachsigen Mercatorprojektion, Google Maps verwendet eine normale Mercatorprojektion. Die Abweichungen der beiden Projektionen sind bereits auf dem relativ kleinen Gebiet des Kantons Zürich derart gross, dass eine befriedigende Überlagerung unserer Karte (in Form einer Rasterbilddatei) mit einem einfachen Verzerren oder Verschieben nicht möglich war, wie man das hier einfach feststellen kann:

Grafik 12: Passprobleme ohne Neuprojektion der überlagerten Karte

10-Meter-Raster der Waldbedeckung



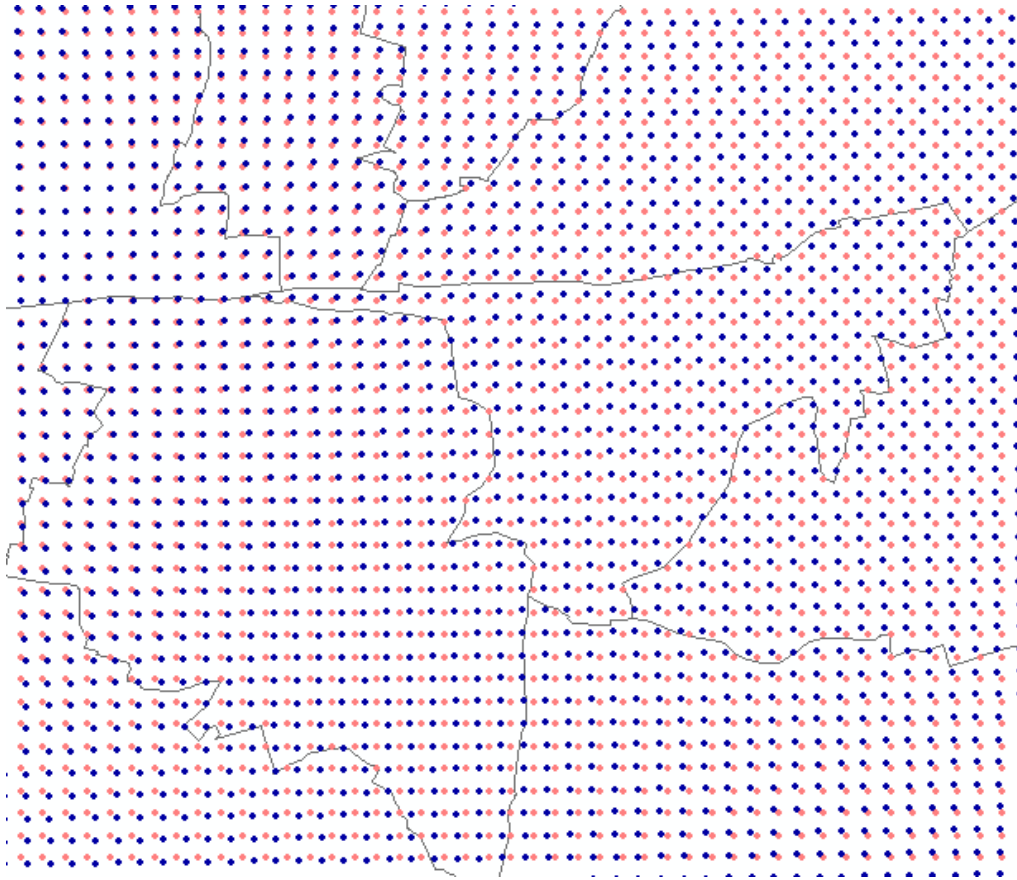
Für die Überlagerung musste deshalb eine saubere – aber auch etwas aufwändigere – Lösung gefunden werden. Dabei wurde von der resultierenden Google Map ausgegangen und quasi ein Resampling der vorhandenen statistischen Informationen vorgenommen. Google Map projiziert das Gebiet auf das regelmässige Bildschirmraster mit einer Mercatorprojektion. Somit können die Koordinaten dieses regelmässigen Rasters auf einfache Weise in Länge/Breite (definiert durch das geodätische Referenzsystem WGS84) zurückgerechnet werden⁹. Damit erhält man ein Raster von Koordinaten mit Länge/Breite-Angaben, welches auf den Bildschirm projiziert zu einem quadratischen Raster wird. Für jeden dieser Punkte wird nun sein nächstgelegener Hektarrasterpunkt gesucht¹⁰ und für diesen die Berechnungen vorgenommen. Der so errechnete Raster passt nun exakt auf den Kartenhintergrund:

⁹ $y = \operatorname{atanh}(\sin(\lambda))$ bzw. $\lambda = \operatorname{atan}(\sinh(y))$. Formel z.B. bei <http://mathworld.wolfram.com/MercatorProjection.html>

¹⁰ Die Projektion der Längen-/Breiten-Werte ins Schweizer Koordinatensystem erfolgen mit der Näherungslösung (Bundesamt für Landestopografie swisstopo 2008, 11). Deren Genauigkeit von einem Meter reicht für die vorliegenden Bedürfnisse problemlos.

Grafik 13: Hektarraster und Bildschirmraster

Für den Wert der Punkte des Bildschirmrasters (blau) wurde jeweils der nächstliegende Punkt des Hektarrasters (rot) verwendet.

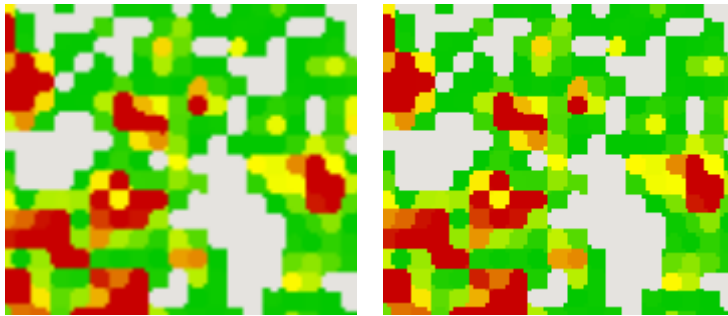


Gut sichtbar wird das Verfahren auf den 500-Meter-Rasterkarten, bei denen dasselbe Verfahren angewendet wurde. Die ursprünglich auf dem schweizerischen Koordinatensystem quadratischen Rasterzellen (bestehend aus 5 mal 5 Hektarzellen) werden auf das 100m-Rasterzellensystem der Google-Karte gezwungen, wobei der Effekt der unterschiedlich projizierten Systeme deutlich wird¹¹:

¹¹ Die beiden Screenshots zeigen im Übrigen gleich auch, wie unterschiedliche Browser (links Firefox 3.5.5 und Internet Explorer 7, je unter Windows XP) die Raster (in Form von PNG-Dateien) je etwas unterschiedlich über die Karte zeichnen.

Grafik 14: Resampling des 500-Meter-Rasters

Beispiel mit zwei verschiedenen Web-Browsern

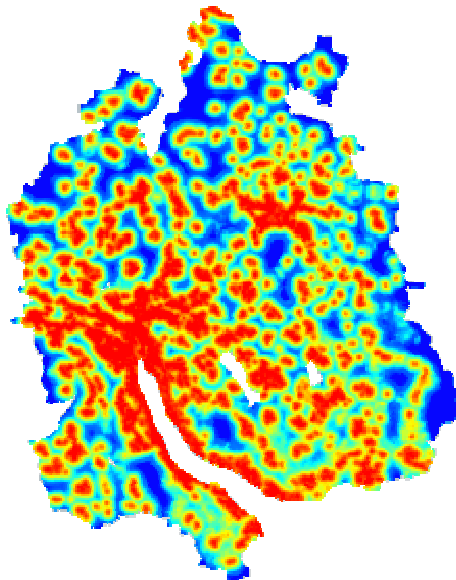


Mögliche Varianten der Gebietsabdeckung

In einem gewissen Sinne stehen die hier erarbeiteten Karten in Verwandtschaft zu dasy-metrischen Karten (Eicher, Brewer, 2001). Bei diesem interessanten, aber schwierig umzusetzenden und deshalb auch selten verwendeten Kartentyp wird ebenfalls ver-sucht, die üblichen Gebietsabgrenzungen zu überwinden und ad-hoc-Gebiete mit gleichartiger Variablenausprägung auszuscheiden. Dazu wird in der Regel auf weitere Informationen (z.B. die Landnutzung) abgestützt. Dies ist bei der vorliegenden Metho-de nicht nötig, trotzdem wäre zu erwägen, in Gebieten, wo die Information für die k nächsten Nachbarn in einer (zu) weiten Umgebung zusammengesucht werden muss, auf eine Visualisierung zu verzichten.

Grafik 15: Informationsdichte

Distanz zu den nächsten 200 Personen (2008)

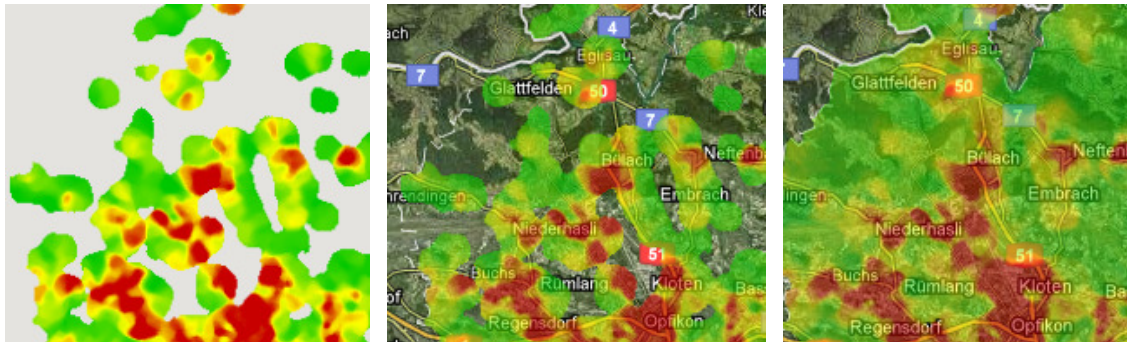


Ausgehend von der Karte der Distanz zu den nächsten 200 Personen wäre zu überle-gen, ob diejenigen Gebiete des Kantons, bei denen es einen grossen Umkreis braucht, um diese Zahl von Personen zu finden (in den blauen, hellblauen, grünlich oder gar gelblichen Gebieten), bei der Darstellung ausgeschlossen werden sollten. Die Visualisie-

rung würde also auf diejenigen Gebiete beschränkt, bei denen eine definierte minimale Informationsdichte erreicht wird.

Grafik 16: Beschränkung auf die Darstellung der besiedelten Gebiete

Ausländeranteil



Während beim Verzicht auf einen Hintergrund diese Variante noch überzeugen kann, zerfällt die Sache bei der Projektion auf ein Satellitenbild: die Ablenkung durch den Hintergrund macht es zu schwierig, die statistische Oberfläche als Ganzes wahrzunehmen, allfällige grössere Strukturen zerfallen weitgehend.

Zur zusätzlichen Verdeutlichung der unterschiedlichen Informationsdichte sind in der Kartographie noch keine allgemein anerkannten Methoden vorhanden (vgl. MacEachren et al. 2005). Man könnte sich allenfalls vorstellen, die Transparenz der Karten oder die Sättigung der Farben von der Informationsdichte abhängig zu machen. Ob sich derartige Methoden eignen würden, müsste im vorliegenden Fall noch abgeklärt werden. Das Problem mit der Interferenz durch die Hintergrundkarte würde aber leider im Falle der Transparenz weiterbestehen.

Aufgrund der gemachten Überlegungen haben wir deshalb am Schluss auf eine spezielle Behandlung der Gebiete mit geringerer Informationsdichte verzichtet und färben in der Regel das ganze Gebiet des Kantons Zürich ein. Dies ermöglicht einen möglichst guten Gesamteindruck über die räumliche Ausprägung der Variablen.

5. Spezielles

Im vorliegenden online-Strukturatlas verwenden wir somit eine neuartige Generalisierungsmethodik für statistische Oberflächen, insbesondere für rasterbasierte Daten (hier Hektaren). Bei üblichen Choroplethenkarten wird die darzustellende Information innerhalb eines Gebietes, welches durch ein Polygon definiert wird, aggregiert und dieses Polygon dann entsprechend dargestellt. Auch eine einfache Rasterkarte entspricht diesem Konzept, allerdings sind dort die Polygone dort sehr klein und quadratisch. Normalerweise sind Visualisierungseinheit und Raumbezug der Objekte also ein und dasselbe, der Raumbezug ist direkt dargestellt und erkennbar. Mit dem Konzept der Aggregation auf Nachbarschaften ist der Raumbezug nicht mehr direkt in der Karte sichtbar. Dies ist gewohnungsbedürftig, ermöglicht aber neue Einsichten, indem jeder einzelne Punkt der Karte adaptiv die spezifische Information seiner Nachbarschaft erhält.

Die hier entwickelte Methode stellt in einem gewissen Sinne eine in Bezug auf den Raumdimension konsequente Weiterentwicklung des Konzepts der nicht klassierten Choroplethenkarte: während bei diesem Kartentyp darauf verzichtet wird, die Datenwerte in – mit welcher ausgeklügelten Methoden auch immer definierte – Klassen zu

zwängen, löst sich bei den hier erarbeiteten Karten auch der mit unstetigen Grenzen behaftete Raumbezug auf und macht kontinuierlichen Übergängen Platz.

Die hier erarbeiteten Karten stellen so einen neuartigen Kartentyp «Umgebungs-» bzw. «Nachbarschaftskarte» dar: sie zeichnen sich durch einen speziellen, adaptiven Raumbezug aus, bei dem die räumliche «Nachbarschaft» statt willkürlich festgelegter Gebietsabgrenzungen (Polygone oder Raster) die Basis für die Visualisierung bildet.

Durch die statistisch einheitliche Informationsdichte werden auch die Belange des Datenschutzes gewährleistet, welcher bei konventionellen Hektarrastern, wo im Extremfall einzelne Personen die Variablenausprägung ausmachen können, auftreten kann. Insofern wird damit eine für thematische Karten fast ideale Generalisierung erreicht: die dargestellten Werte sind für jede Teilregion gleich zuverlässig. Die Umgebungsmethode eliminiert so das «Rauschen» in konventionellen Hektarrastern und bringt die Hauptstrukturen zum Vorschein.

6. Ausblick

Die hier erarbeitete Methodik bildet die Basis für die weitere Erkundung des Raums «Kanton Zürich». Unter den möglichen Weiterentwicklungen und Anwendungen stehen Bereiche zur Auswahl: in thematischer Hinsicht werden wir weitere Ebenen hinzufügen. Mögliche Beispiele sind etwa die Tagesbevölkerung, Ergebnisse von Modellrechnungen (z.B. Bodenpreise) und andere Themen (z.B. die Daten der Motorfahrzeugstatistik). In analytischer Hinsicht wäre es interessant, ein paar zusätzliche Kennzahlen zur statistischen Charakterisierung der Oberflächen hinzuzufügen (z.B. die jeweilige räumliche Autokorrelation).

Die Fülle der Daten legt es im Weiteren nahe, auch faktoranalytische Methoden oder Clusteranalysen einzusetzen, um aus der Vielfalt der Variablen und deren räumlichen Ausprägungen die grundlegenden Dimensionen bzw. Teilräume zu erarbeiten.

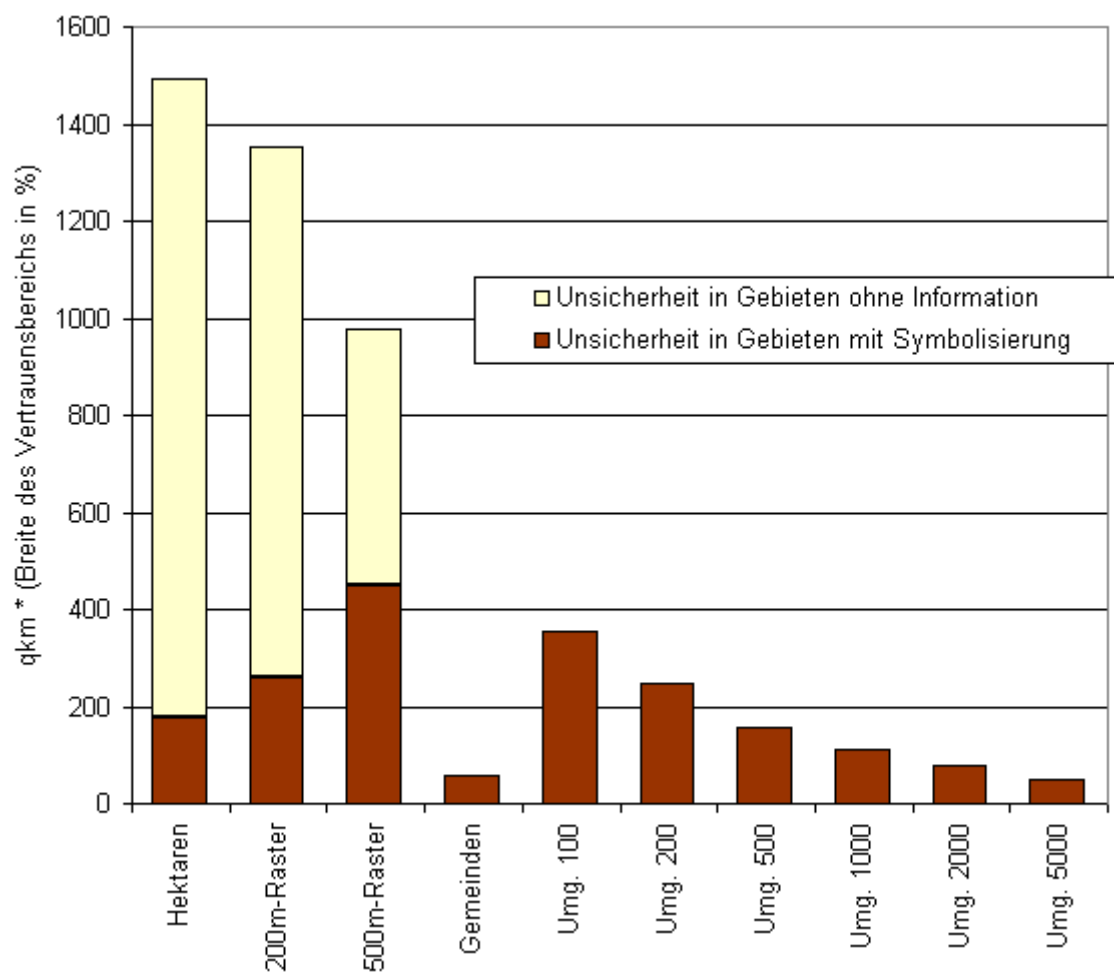
Anhang: Statistische Zuverlässigkeit der Methoden

Schätzung der gesamten Unsicherheit einer Karte

Zur Schätzung der gesamten Unsicherheit in einer Karte haben wir als Mass die Aufsummierten Produkte von Fläche und Unsicherheit berechnet, welche wir hier noch kurz zeigen möchten. Interessant ist zunächst der Vergleich der verschiedenen Raster (vgl. Grafik 17): während beim Hektarraster ein grosser Teil des Kantons überhaupt keine Werte aufweist, ist die in den besiedelten Gebieten auflaufende Menge an Unsicherheit im Verhältnis eher klein. Bei einer Vergrößerung des Rasters nimmt der Teil des Kantons ohne jede Information ab (und damit auch die gesamte Unsicherheit; Verkleinerung des gelben Teil des Balkens), allerdings wächst die Unsicherheit innerhalb des kartierten Gebietes: die zusätzlich dargestellten Gebiete (v.a. am Rande der Siedlungen) zeigen Informationen, die auf nur kleinen Zahlen beruhen und deshalb die Gesamtunsicherheit zusätzlich erhöhen (brauner Teil der Balken). Die Umgebungskarten schneiden dagegen gut ab: sie weisen auch bei räumlich detaillierter Information nur eine geringe Menge an Unsicherheit auf.

Grafik 17: Menge an Unsicherheit in verschiedenen Karten

Beispiel mit einem angenommenen Prozentwert von 20; Flächengewichtung



Literaturverzeichnis

- BBC: *The Truth About Crime – The Oxford Crime Map*. URL <http://www.bbc.co.uk/truthaboutcrime/crimemap> (8.10.2009)
- BIVAND, Roger S.; PEBESMA, Edzer J. ; GÓMEZ-RUBIO, Virgilio (2008): *Applied Spatial Data Analysis with R*; Springer
- BUNDESAMT FÜR LANDESTOPOGRAFIE SWISSTOPO (2008): *Formeln und Konstanten für die Berechnung der Schweizerischen schiefachsigen Zylinderprojektion und der Transformation zwischen Koordinatensystemen*, Oktober 2008, Bern
- EICHER, Cory L. ; BREWER, Cynthia A. (2001): Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. In : *Cartography and Geographic Information Science*, April 2001
- GETIS, A. ; ORD, J. K. 1996 Local spatial statistics: an overview. In: Longley P. ; Batty, M. (eds): *Spatial analysis: modelling in a GIS environment*, Cambridge: Geoinformation International), 261–277.
- GOOGLE: *Google Maps-API – Referenz*, URL <http://code.google.com/intl/de-DE/apis/maps/documentation/reference.html> (3.11.09)
- HAKE, Günter; GRÜNREICH, Dietmar; MENG, Liqiu (2002): *Kartographie*, 8. Auflage Berlin; New York : de Gruyter
- HERZOG, Adrian ; HOFSTETTER, Claudia (2009): *Die Umgebungsanalyse des Statistischen Amts. Version 2.0*, Zürich : statistik.info 03/2009
- IMHOF, Eduard (1972): *Thematische Kartographie*, Lehrbuch der Allgemeinen Geographie, Band X, Berlin, New York: Walter de Gruyter.
- MAC EACHREN, Alan M ; ROBINSON A. ; HOPPER S. ; GARDNER St. ; MURRAY R. ; GAHEGAN M. ; HETZLER E. (2005): Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. In: *Cartography and Geographic Information Science*, Vol. 32, No. 3, pp. 139-160
- OPENSHAW, Stan ; TAYLOR, P.J. (1981): The Modifiable Areal Unit Problem. In: WRIGLEY, Neil ; BENNET, R.J. (eds.), *Quantitative Geography: A British View*, London, 60-69
- ORD, J. K. ; GETIS, A. (1995): Local spatial autocorrelation statistics: distributional issues and an application. In: *Geographical Analysis*, 27, 286–306;
- SCHULER, Martin; BOPP, Matthias ; BRASSEL, Kurt E. ; BRUGGER, Ernst A. (Hrsg., 1985): *Strukturatlas der Schweiz, Atlas structurel de la Suisse*, Zürich: Ex Libris
- SCHULER, Martin ; DESSEMONTET, Pierre ; et al. (2006): *Atlas des räumlichen Wandels der Schweiz*, Bundesamt für Statistik (BFS), Neuchâtel, Zürich: Verlag Neue Zürcher Zeitung.
- SCOTT, D.W. ; SAIN, S.R.(2004): Multi-Dimensional Density Estimation. In: RAO, C.R. ; WEGMAN, E.J. (eds.): *Handbook of Statistics, Vol 23: Data Mining and Computational Statistics*, Amsterdam: Elsevier.
- WOLFRAM MATHWORLD: Mercator Projection, URL <http://mathworld.wolfram.com/MercatorProjection.html> (3.11.09).

Das Statistische Amt des Kantons Zürich ist das Kompetenzzentrum für Datenanalyse der kantonalen Verwaltung. In unserer Online-Publikationsreihe "statistik.info" analysieren wir für ein breites interessiertes Publikum wesentliche soziale und wirtschaftliche Entwicklungen im Kanton und Wirtschaftsraum Zürich.

Unser monatlicher Newsletter "NewsStat" und unser tagesaktueller RSS-Feed informieren Sie über unsere Neuerscheinungen in der Reihe "statistik.info" sowie über die Neuigkeiten in unserem Webangebot.

Fragen, Anregungen, Kritik?

Adrian Herzog und Claudia Hofstetter

Telefon: 044 225 12 55 / 044 225 12 16

E-Mail: adrian.herzog@statistik.ji.zh.ch / claudia.hofstetter@statistik.ji.zh.ch

Statistisches Amt des Kantons Zürich

Bleicherweg 5

8090 Zürich

Telefon: 044 225 12 00

Fax: 044 225 12 99

E-Mail: datashop@statistik.zh.ch

www.statistik.zh.ch

© 2009 Statistisches Amt des Kantons Zürich, Abdruck mit Quellenangabe erlaubt