



Kanton Zürich
Direktion der Justiz und des Innern
Statistisches Amt

One tool to rule them all - R als statistisches Sackmesser

Dr. Peter Moser

R - ein multifunktionales Statistikwerkzeug

Im STAT seit 2000 verwendet, anfangs vor allem in der Analyseabteilung, heute amtsweit im Einsatz

- Warum R?
 - Programmiersprache – fördert strukturiertes Denken in Funktionen und Abläufen und die Replizierbarkeit von Auswertungen und Analysen
 - Weltweite Nutzergemeinschaft - Lösungen für (fast alle) denkbaren Probleme sind im Internet zu finden
 - Open source - getragen von einer breiten Entwicklercommunity, kein kostspieliges Lizenzmodell
 - Erweiterbarer Programmkernel - Funktionalität ist durch "libraries", Codebibliotheken ausbaubar (vorteilhaft in einem restriktiven Verwaltungs-IT-Umfeld!)

Vielseitigkeit - In einer Umgebung kann eine Vielzahl analytischer Probleme gelöst werden

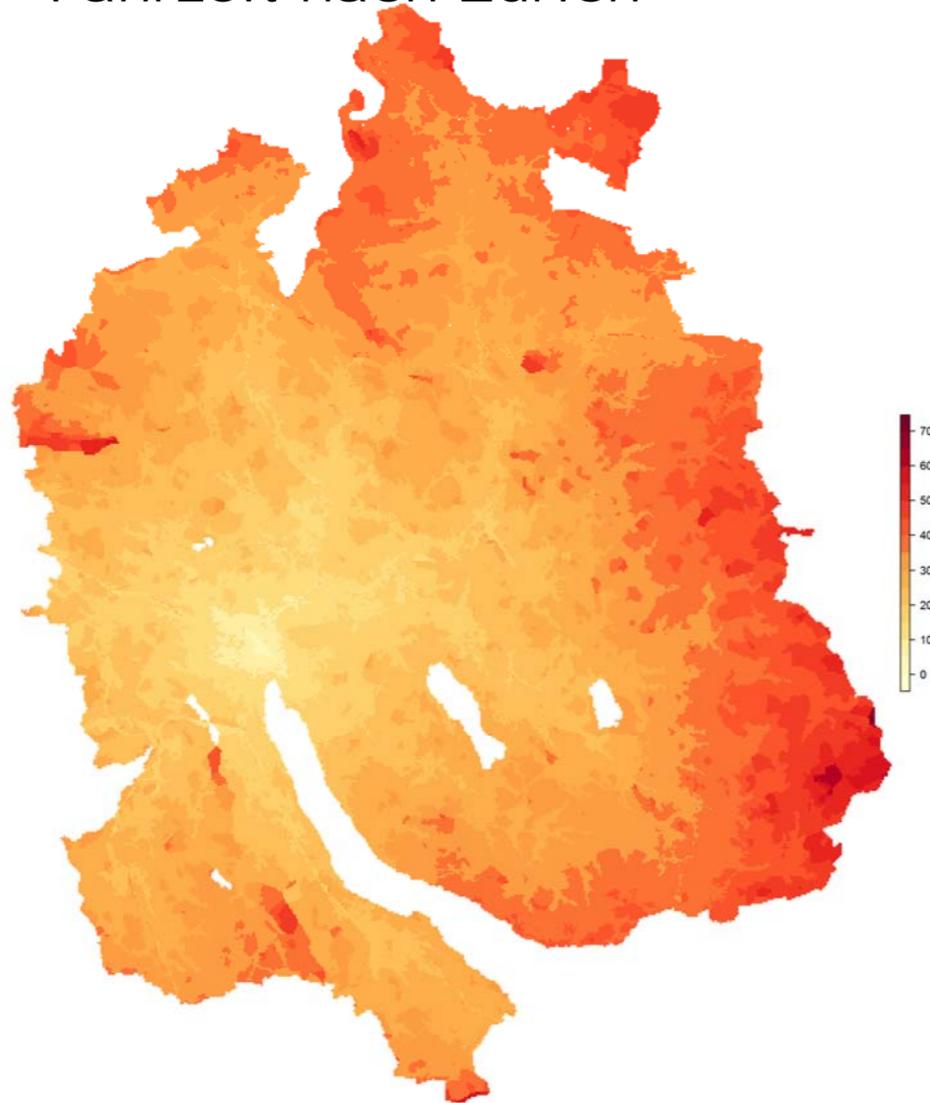
Exempel: ein hedonisches Bodenpreismodell

- Veredelung der notariellen Handänderungsdaten, einer der wertvollsten Datensätze des Amts
- Ziele:
 - Erkenntnisgewinn: Wie funktioniert der Markt für Wohnbauland im Kanton Zürich, Welche Lage- und Grundstückseigenschaften beeinflussen die Preise in welchem Ausmass?
 - Praktische Anwendung: Erzeugung kleinräumiger Schätzwerte für den ganzen Kanton für verschiedene Zwecke
- Methodisch anspruchsvolles, vielfältiges Projekt:
 - Aufbereitung und Berechnung von Mikro und Makro-Lagecharakteristiken aus unterschiedlichsten Quellen
 - Datenmanagement
 - Modellierung
 - Vermittlung und Visualisierung der Resultate

Von (fast) A bis Z mit R umgesetzt

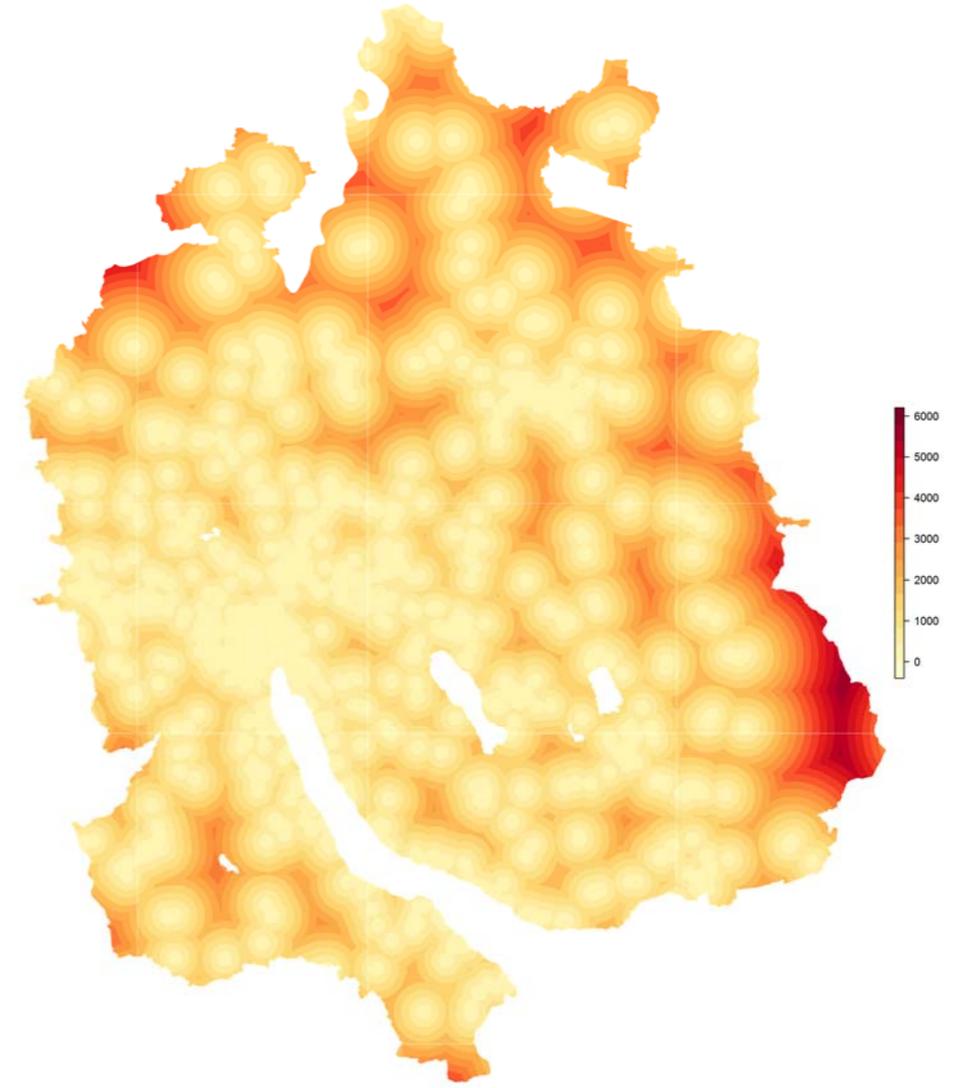
Lageeigenschaften I

Fahrzeit nach Zürich



Grundlage: Google-Routing API

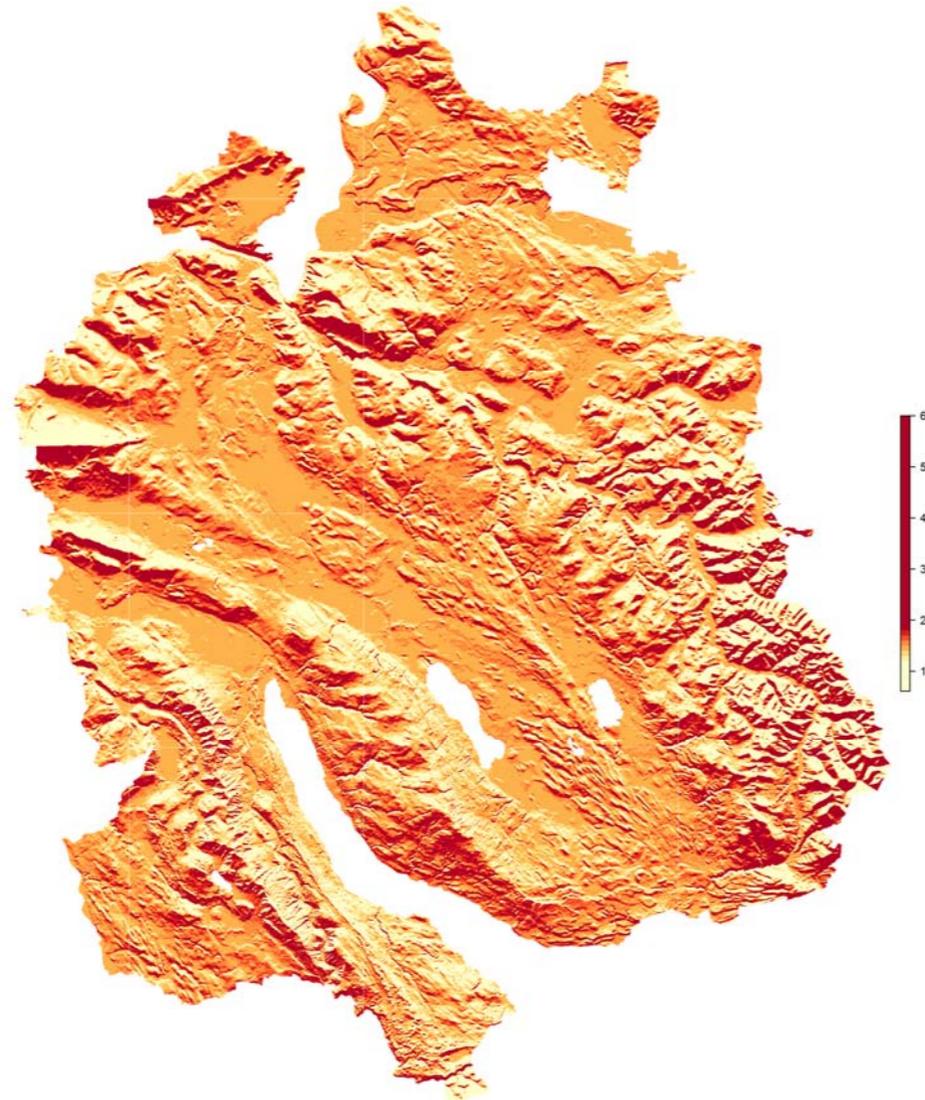
Distanz zu Detailhandelsgeschäften



Grundlage: STATENT

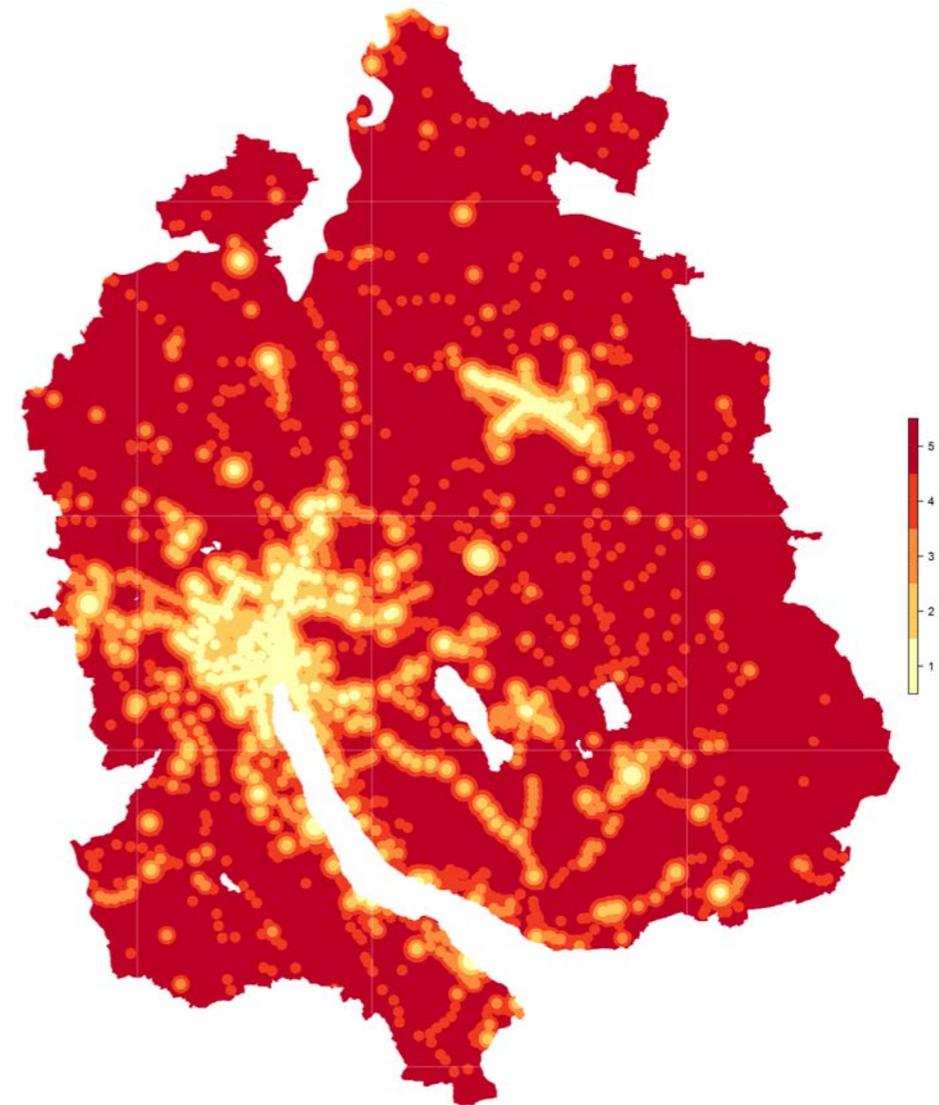
Lageeigenschaften II

Sonneneinstrahlung



Grundlage: DHM25

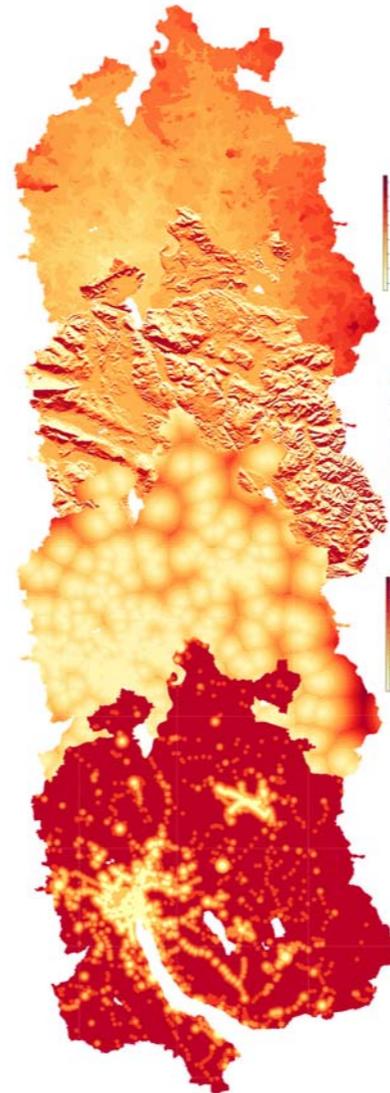
ÖV-Güteklassen



Grundlage: OGD, ARE-CH

Datenaufbereitung

- Lageeigenschaften in 25X25 Meter-Raster: für Wohnbauland im Kanton ZH ~325K Zellen
- 7000 Landtransaktionen: Datenbereinigung und Aufbereitung in Oracle-DB, Lageeigenschaften aus dem Raster
- Transaktionsdaten können dank einheitlicher Projektionsinformationen mit Lageeigenschaften verknüpft werden.



Transaktionen

x	y	gmiv	detailhandel	insol	oevg	qmpreis
697774	232951	39.6	0.3	1.4	3	358.0
687840	233933	23.6	0.4	0.9	4	1382.6
702102	241644	27.3	0.3	1.3	1	328.9
683831	264093	24.5	0.4	1.7	4	1608.4
693364	258787	26.0	0.2	1.6	4	570.0
704212	268992	33.4	2.0	1.5	4	1384.0

Raster

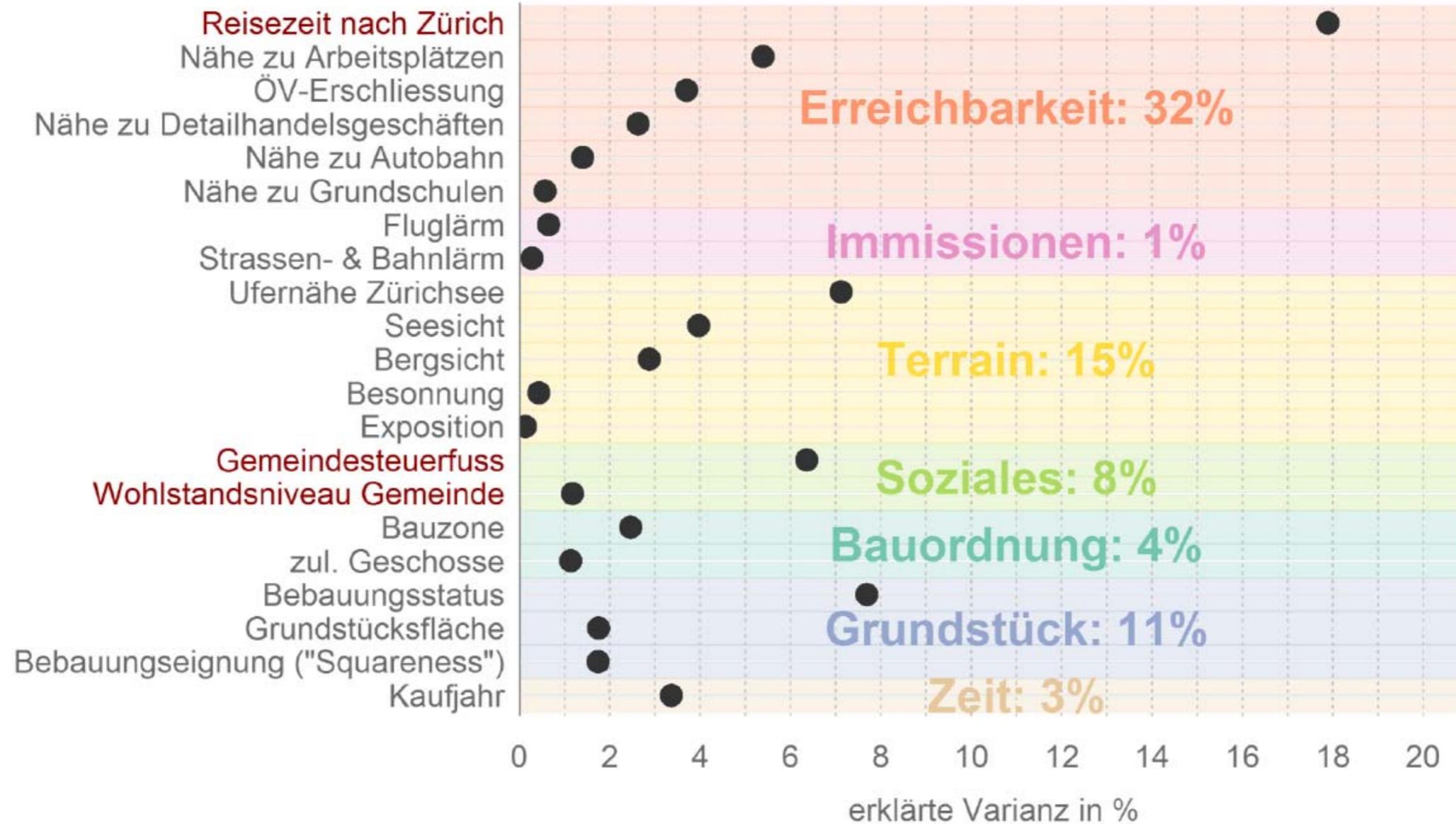
x	y	gmiv	detailhandel	insol	oevg
690213	283288	39.6	0.2	1.2	2
690238	283288	39.6	0.2	1.2	2
690263	283288	39.6	0.2	1.3	2
690288	283288	39.6	0.2	1.3	2
690313	283288	39.3	0.2	1.4	2
690338	283288	39.3	0.2	1.4	2

Modellieren - die "Kernkompetenz" von R

- Die ganze Bandbreite moderner statistischer Modellierungstechniken steht zur Verfügung:
- Zur Modellierung der "lärmigen" Bodenpreise wird ein robustes Modell verwendet.
- Output auf die Konsole ist möglich -
- In der R-Programmiersprache ist das Modell aber ein Objekt, das alle nötigen Informationen enthält (Parameter, Residuen, robuste Gewichte, design matrix etc.) und so seinerseits als Input dient.

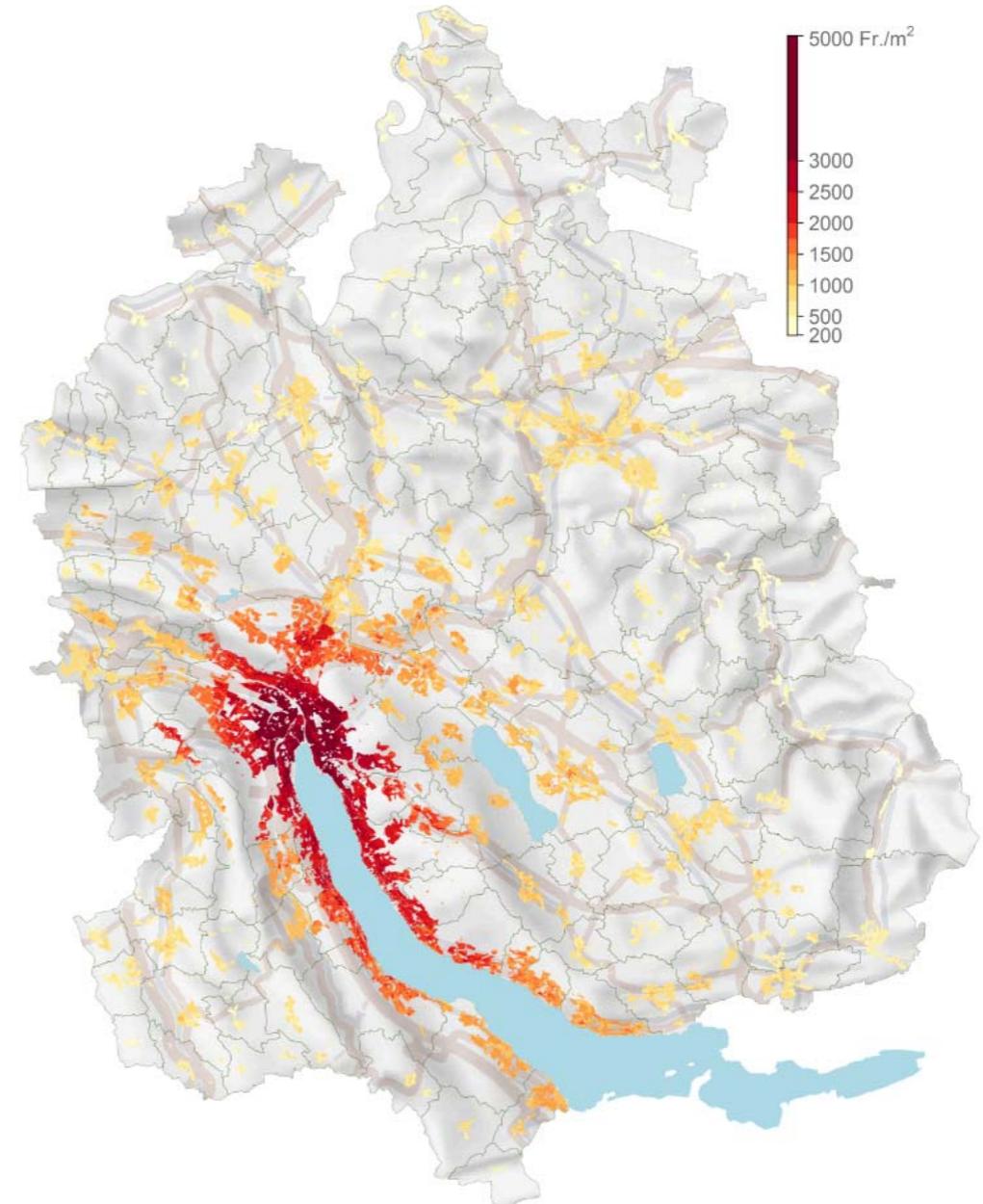
	log(qmpreis)
log(mean_kern_mivoev)	-0.495*** (-0.529, -0.461)
distarb200ha	-0.027*** (-0.035, -0.019)
distdetailhandel	-0.072*** (-0.089, -0.054)
distprimarschulen	-0.094*** (-0.141, -0.047)
oevgkarech_n	-0.020*** (-0.033, -0.007)
hlstrassendist	0.023*** (0.009, 0.037)
I(hlstrassendist2)	-0.002*** (-0.004, -0.001)
laerm_strasse_bahn	-0.016*** (-0.022, -0.010)
fluglaerm_n_t_db	-0.043*** (-0.052, -0.034)
sin(aspect)	-0.039*** (-0.055, -0.022)
insol2	0.093*** (0.045, 0.141)
log(bergs + 1)	0.099*** (0.079, 0.118)
log(seesicht + 1)	0.058*** (0.022, 0.095)
seedist_trunc	0.071*** (0.054, 0.089)

Resultate I: Bedeutung der Einflussfaktoren

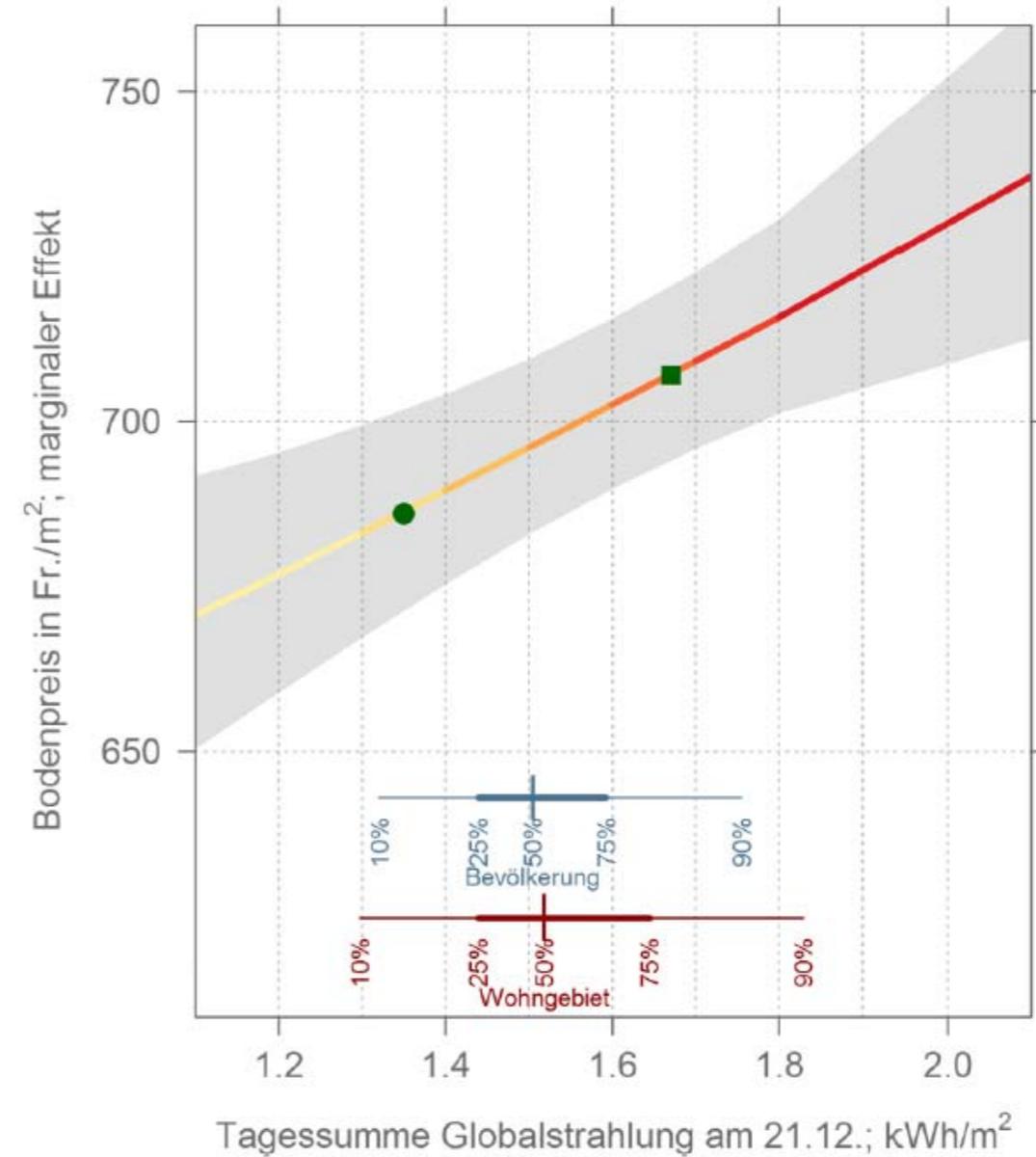
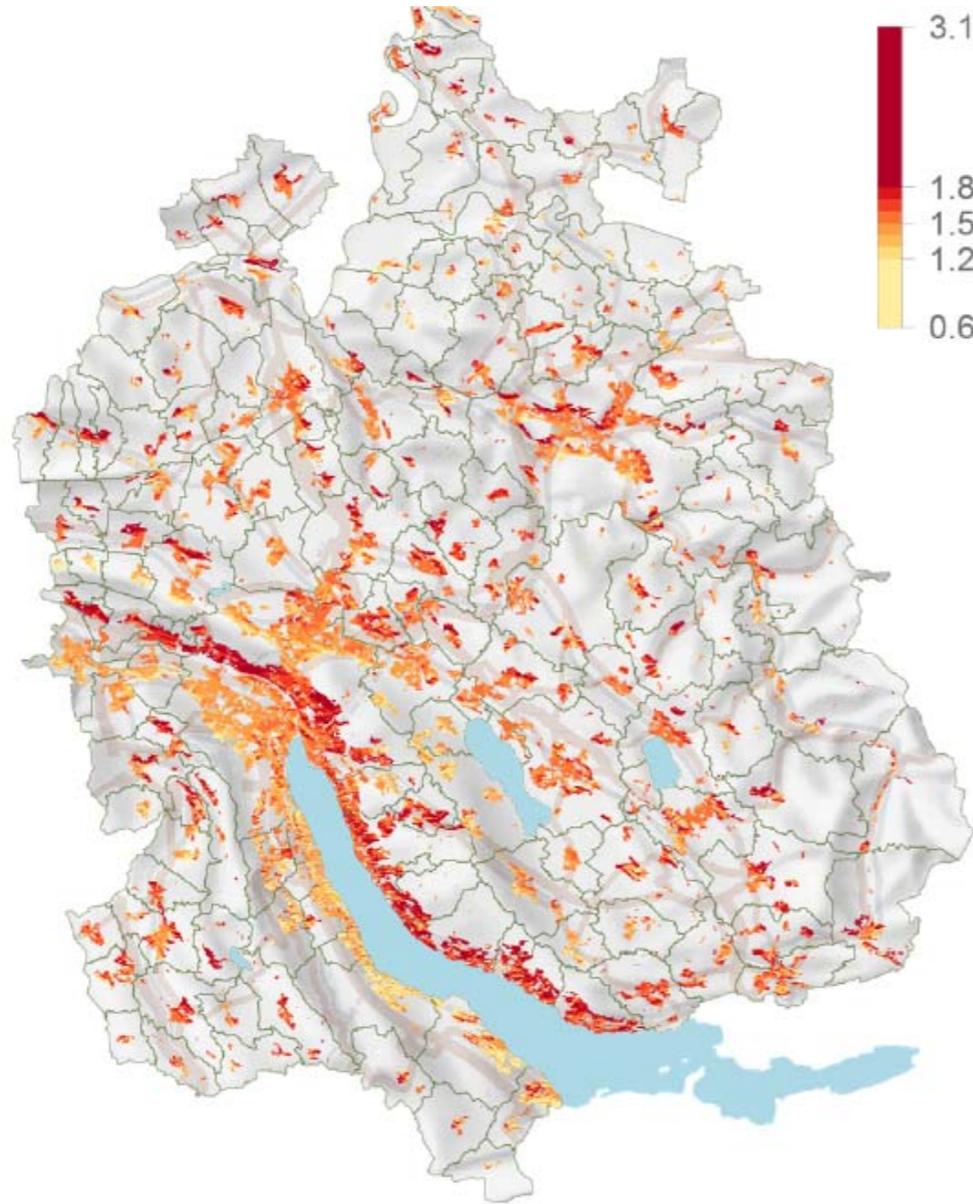


Resultate II: kleinräumige Schätzwerte

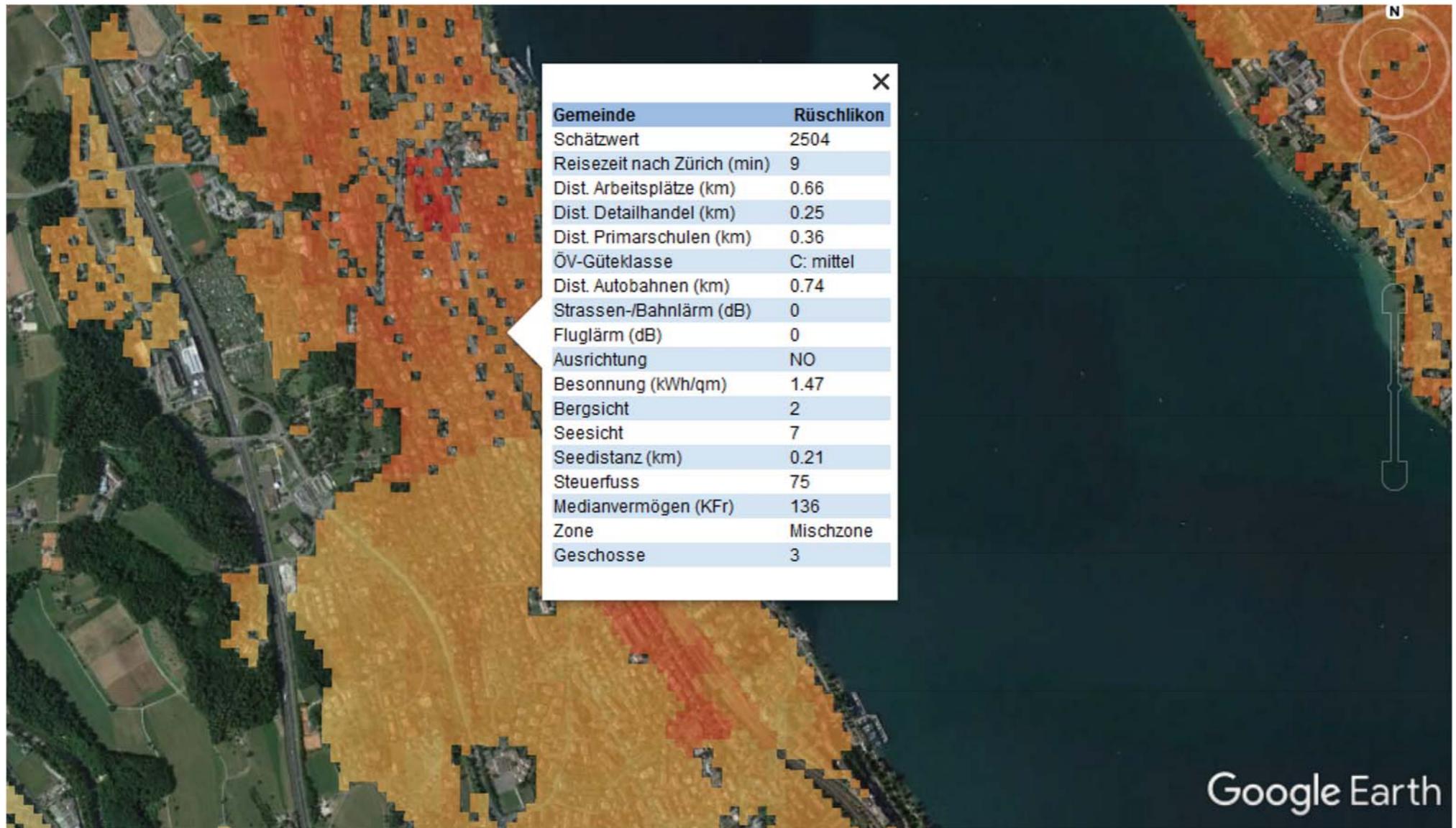
- Die Schätzwerte werden durch Kombination des Modellobjekts mit den Lageeigenschaften im Rasterobjekt erzeugt.
- Reliefschattierung, wichtige Verkehrswege sind bei Swisstopo als GEOtiff (Bildformat mit Koordinaten und Projektionsinformationen) verfügbar.
- Gemeindegrenzen (ArcGis Shapefile)
- können deshalb problemlos kombiniert werden



Resultate III: Besonnung – Verteilung und Effekt



Resultate IV: Schätzwerte in .kml (Google-Earth)



Weiterführende Informationen

Verwendete R-libraries:

- Aufbereitung, Handling, Verarbeitung räumlicher Daten: raster, sp, rgdal, maptools, rgeos, cleangeo, insol....
- Modellierung: robustbase, effects, relimp, car, stargazer, FNN....
- Visualisierung: lattice, latticeExtra, gridExtra, leafletR, plotKML, classInt, diagrammeR ...

Die Publikation:

"Der Preis des Bodens – Ein hedonisches Modell der Landpreise im Kanton Zürich
statistik.info 05/2017

Koordinaten:

peter.moser@statistik.ji.zh.ch

Twitter: [@statistik_zh](https://twitter.com/statistik_zh); [@peterjamoser](https://twitter.com/peterjamoser)