



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 2000 eingeführt, seither allmähliche Verbreitung, anfangs primär in der Analyseabteilung, heute amtsweit im Einsatz
- Warum R?
 - Objektorientierte Programmiersprache – jeder Output ist auch potentieller Input – diszipliniert und fördert Replizierbarkeit von Auswertungen und Analysen
 - Ein Programmkern, dessen Funktionalität durch "libraries", Codebibliotheken erweiterbar ist (vorteilhaft in einem restriktiven Verwaltungs-IT-Umfeld!)
 - Weltweite Nutzergemeinschaft: Unterstützung; Lösungen für (fast alle) denkbaren Probleme sind im Internet zu finden
 - Open source, getragen von einer breiten Entwicklercommunity, kein teures Lizenzmodell

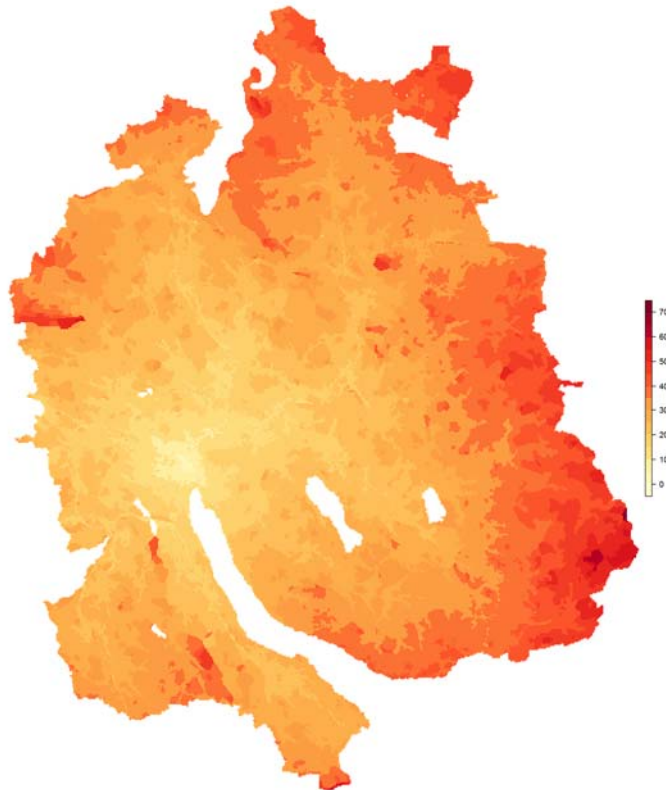
Vielseitigkeit: In einer Umgebung kann eine Vielzahl analytischer Probleme gelöst werden. "One tool to rule them all"

Als Exempel: ein hedonisches Bodenpreismodell

- Veredelung der 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 ins Werk gesetzt

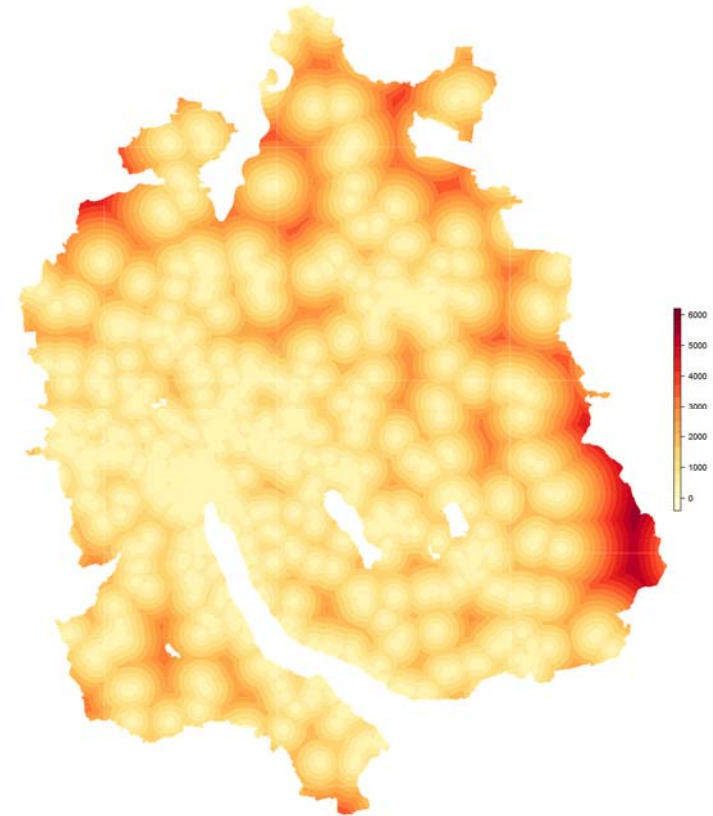
Bsp: Lageeigenschaften I

Fahrzeit nach Zürich



Grundlage: Google-Routing API

Distanz zu Detailhandelsgeschäften



Grundlage: STATENT

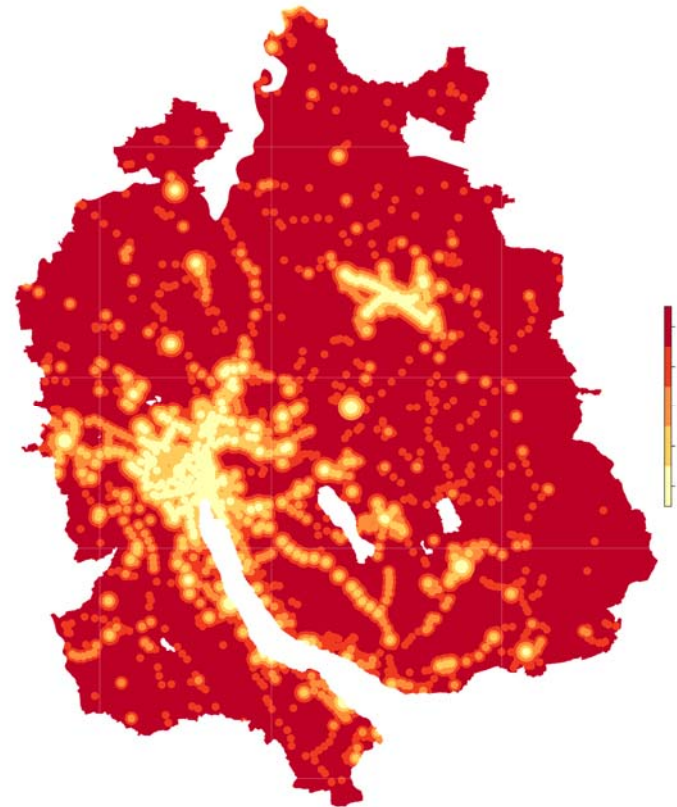
Bsp: Lageeigenschaften II

Sonneneinstrahlung



Grundlage: DTM25

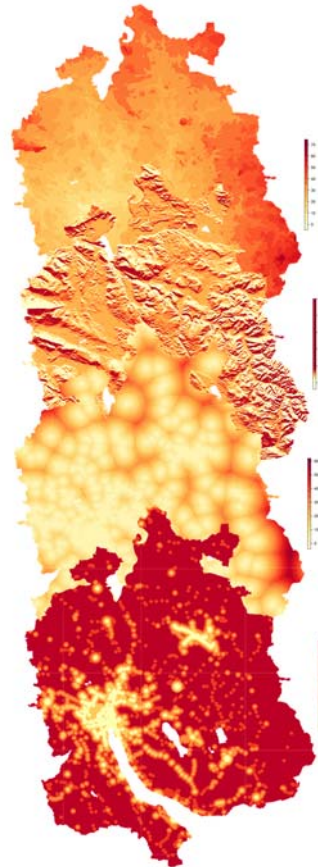
ÖV-Güteklassen



Grundlage: DTM25

Datenbearbeitung

- Lageeigenschaften in 25X25 Meter-Raster (für Wohnbauland ergibt das im Kanton ZH ~325K Zellen)
- Landtransaktionsdaten: Aufbereitung (Bereinigung von Merkmalen etc.) in (Oracle-)DB
- verortete Transaktionsdaten können dank einheitlicher Projektionsinformationen (z. B. CH1903+ LV95) 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 |

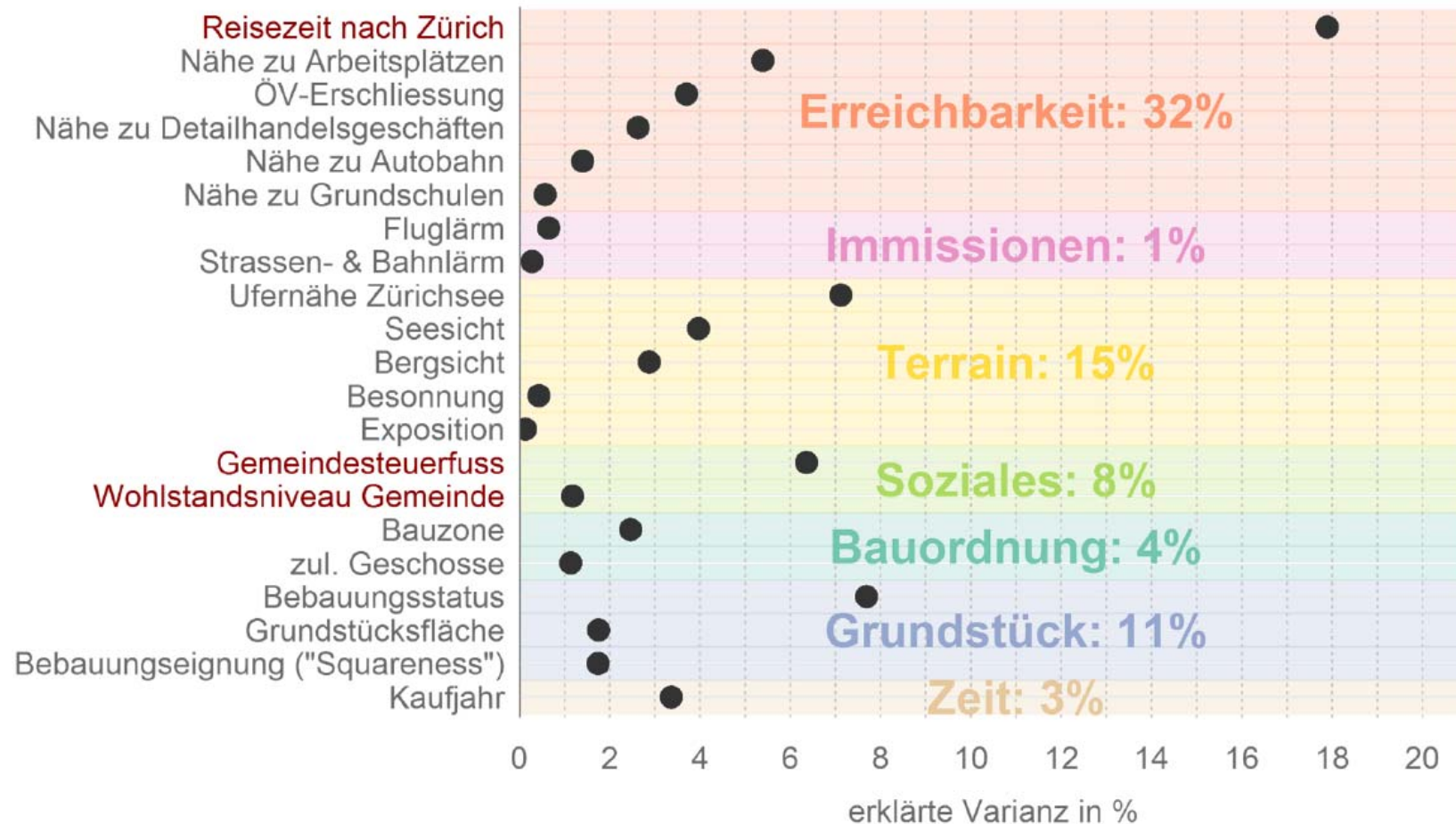
Raster

| 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 |

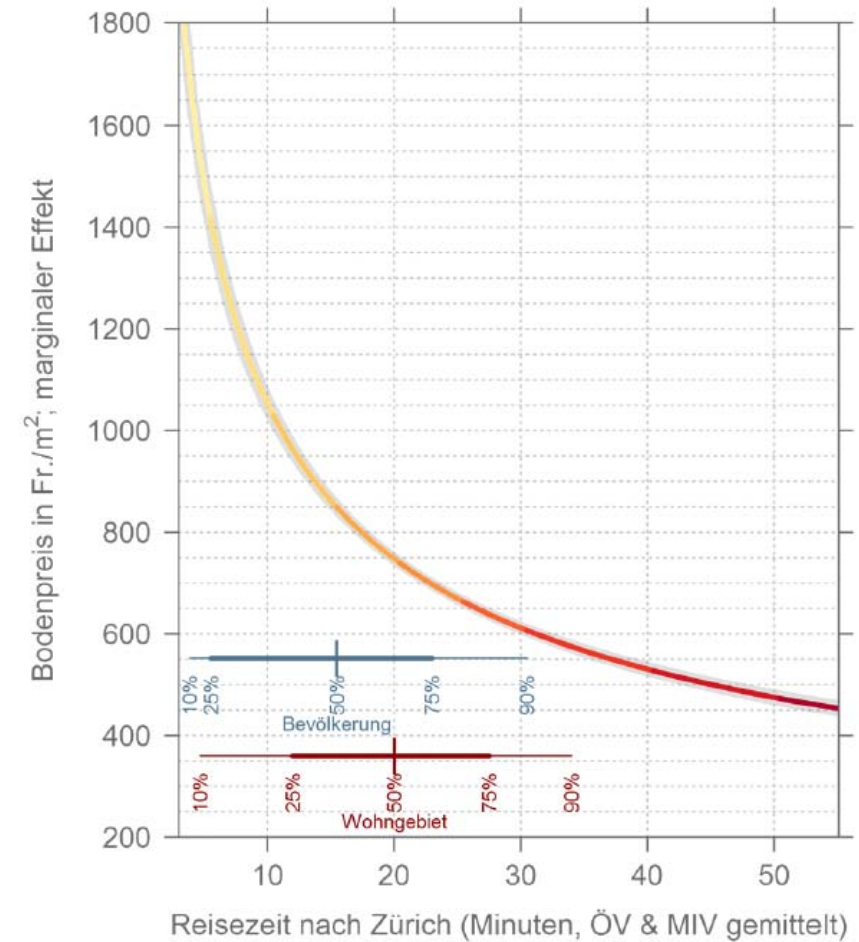
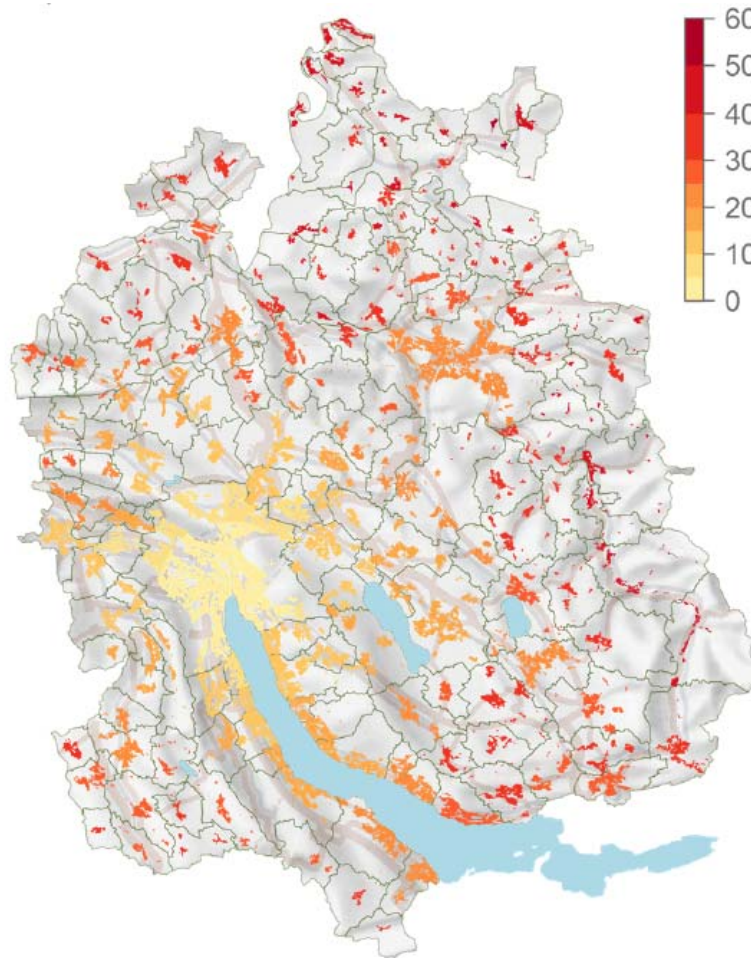
Modellieren - die "Kernkompetenz" von R

- Die ganze Bandbreite moderner statistischer Modellierungstechnologien steht zur Verfügung: Zur Modellierung der "lärmigen" Bodenpreise wird ein robustes Modell verwendet
- In der R-Programmiersprache ist das Modell ein Objekt, das alle nötigen Informationen enthält (Parameter, Residuen, robuste Gewichte, design matrix etc.), kein "Output" auf der Konsole
- Das Modellobjekt ist selbst wieder Input für
 - Modelldiagnose
 - Visualisierungsfunktionen
 - Zusammen mit den Lagecharakteristiken im Raster für die Berechnung von Schätzwerten

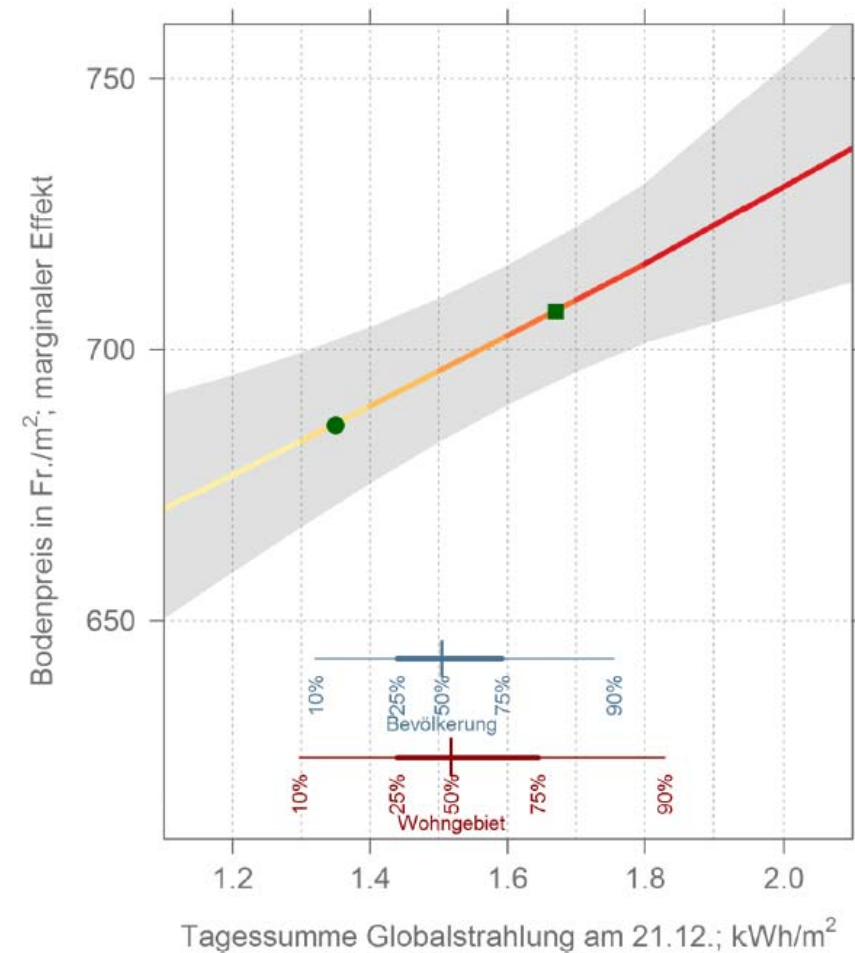
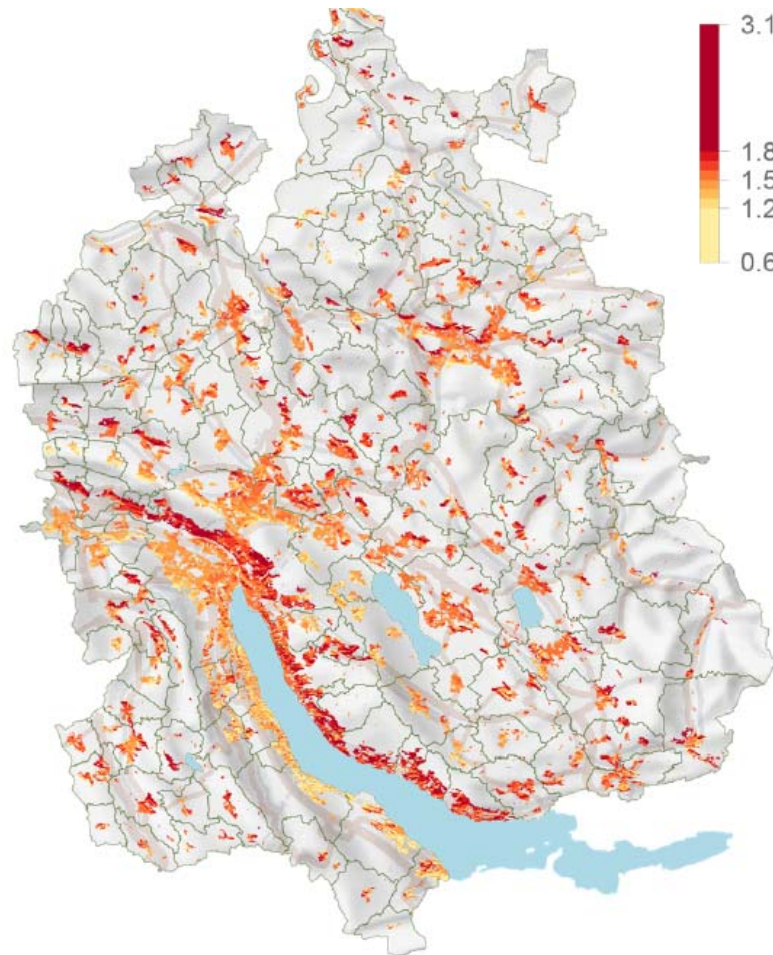
Resultate I: relative Bedeutung der Einflussfaktoren



Resultate II: Reisezeit nach Zürich

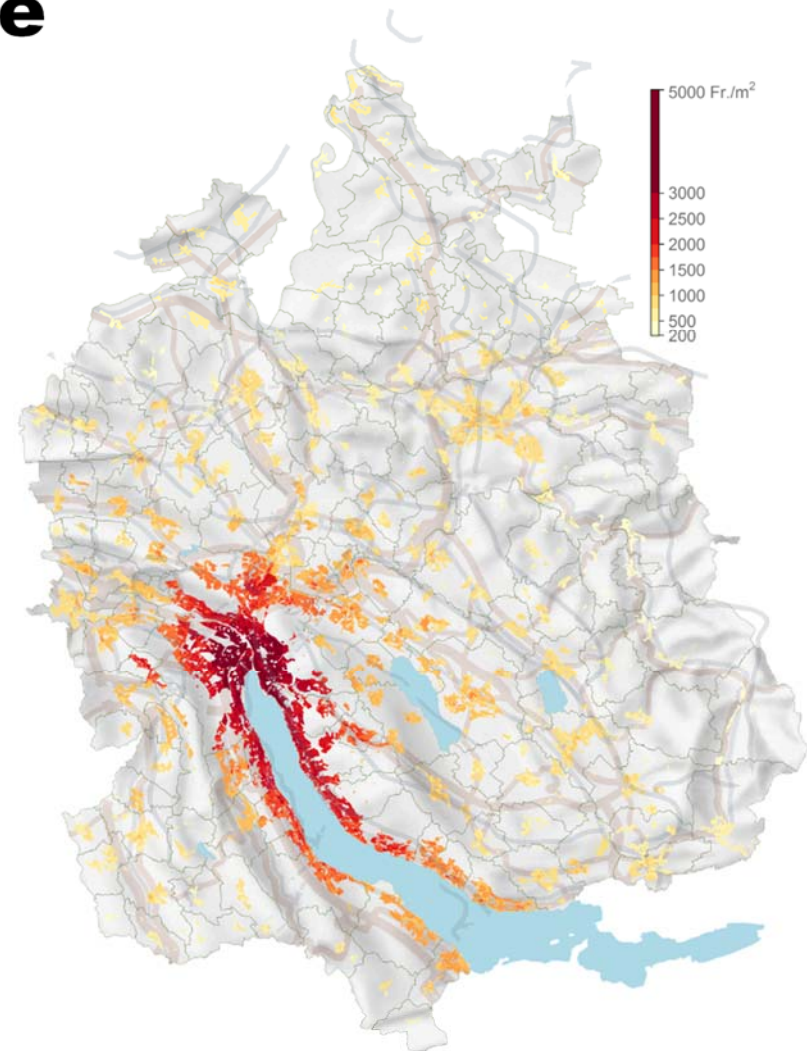


Resultate III: Effekt der Besonnung



Resultate III: Schätzwerte

- Reliefschattierung, Wichtige Verkehrswege von Swisstopo (bereits als geotiff, OGD-verfügbar)
- Auch die Schichten in der Grafik (Relief, Verkehrsinfrastruktur, Gemeindegrenzen,) können kombiniert werden weil sie alle dasselbe LK-Koordinatensystem aufweisen



Export der Schätzwerte in .kml (Google-Earth)



Fazit

- Eine Vielzahl unterschiedlicher Datentypen kann in einem einheitlichen Framework manipuliert werden (neben statistischen Daten, auch räumliche rasterpunkt-, polygon-, Liniendaten)
- Hier nicht von Belang, aber in R können auch Optimierungsprobleme (Operations research, Quadratische Optimierung) gelöst werden - Wahlhochrechnung

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, leafletR, plotKML, gridExtra, classInt...

Fragen?

Dr. Peter Moser

Statistisches Amt des Kantons Zürich

Schöntalstrasse 5

8090 Zürich

peter.moser@statistik.ji.zh.ch

www.statistik.zh.ch

Die resultierende Publikation:

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

Modellieren - die "Kernkompetenz" von R

- Die ganze Bandbreite moderner statistischer Modellierungstechnologien steht zur Verfügung: Zur Modellierung der "lärmigen" Bodenpreise wird ein robustes Modell verwendet
- Das Modell ist kein "Output" sondern ein Objekt, das alle nötigen Informationen enthält (Residuen etc.),
- Dient als Grundlage für weitere

