

Data Science – Von Daten zur Information

Göran Kauermann

9. April 2024





Über mich

- Professor für Statistik an der Ludwig-Maximilians-Universität München
- DFG Fachkollegiat für Statistik und Ökonometrie
- Vorsitzender der German Data Science Society (e.V.)
- Sprecher des Elitestudiengangs Data Science @ LMU
- Vorstand Munich School for Data Science (MUDS)
- Mitglied Munich Center for Machine Learning (MCML)

2 → Forschungsgebiet: Angewandte Statistik



Munich Center for Machine Learning



Inhalt

Themen des Vortrags

1. Daten = Information
2. Daten liefern Information: Die statistische Trickkiste
3. Daten und Datenschutz in Deutschland, Gegner oder Partner?
4. Daten und Qualität: Ein Plädoyer



1. Daten = Information

→ Die europäische Datenschutz-Grundverordnung gilt überall in Europa, die Auslegung ist sehr unterschiedlich

→ **Daten liefern Information für das Gemeinwohl**

→ Es sollte gelten: Gemeinwohl geht vor individuelle Selbstbestimmung

Aber:

→ Es zeigt sich immer wieder, die Digitalisierung (Daten) in Deutschland ist rückständig im Vergleich zu anderen Ländern.

→ Beispiele: Covid-Krise, Kommunale Wärmeplanung, Verkehrsplanung



1. Daten = Information

- ➔ Ein paar Beispiele zum Thema Daten => Information
 - ➔ Wir zeigen, was mit statistischem Instrumentarium möglich ist
1. Bikenetzwerk in der Stadt,
 2. Fahrradnutzung als Kostenersparnis,
 3. Parkplatzsuche in einer Smart City,
 4. Bahnverspätungen anhand von Daten,
 5. Social Media for Good,
 6. Mobilität und Telekommunikation

2. Daten und Information: Die statistische Trickkiste

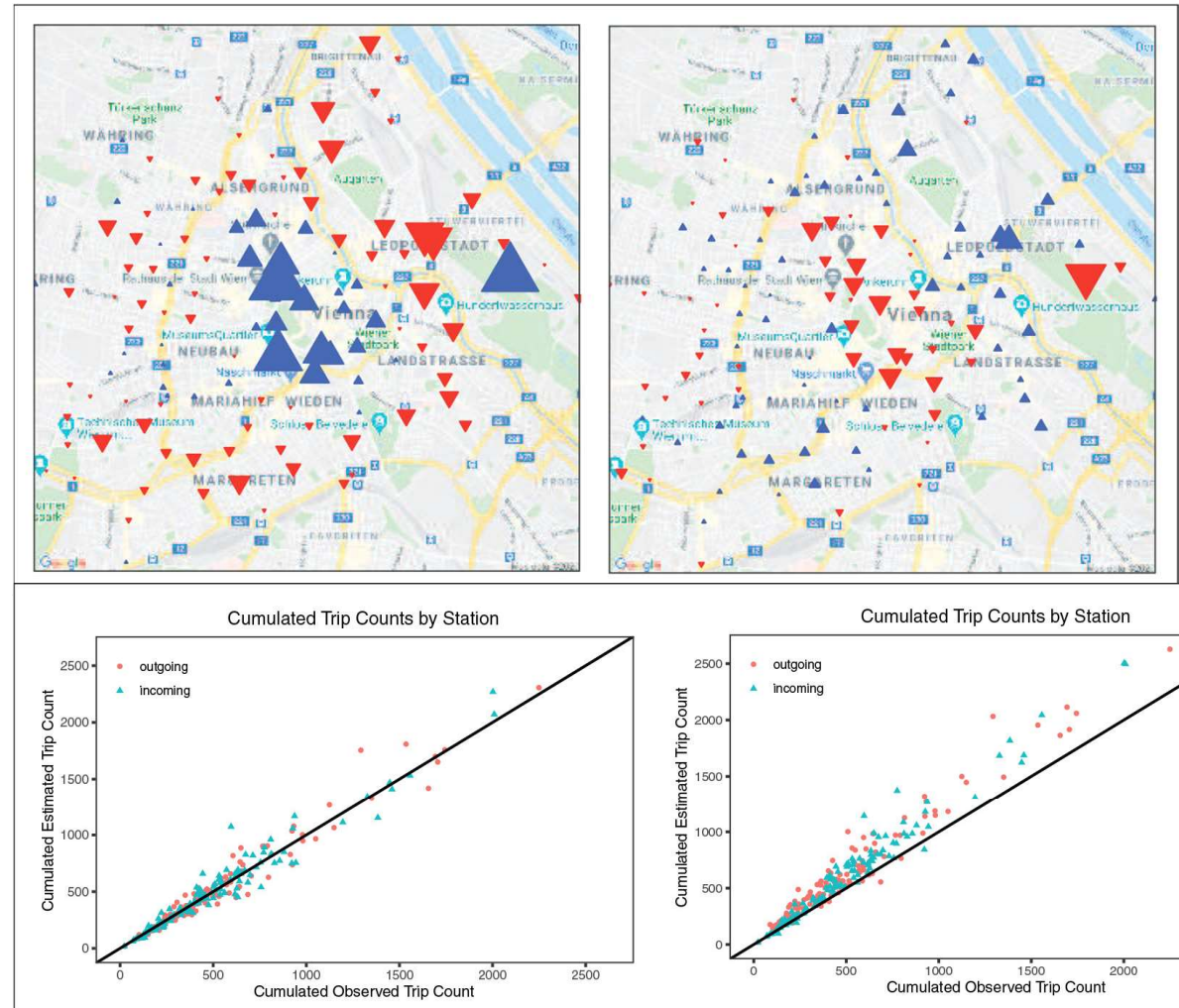
Beispiel 1: Bikenetzwerk

- Bikesharing ist inzwischen in allen größeren Städten verfügbar
- Die Füllstände der Stationen sind (im Prinzip) frei verfügbar.
- Die Daten der Provider geben relevante Information über jede einzelne Fahrt
- Jede Fahrt ist zur Abrechnung verbunden mit einem Nutzer (Stichwort Datenschutz)
- Individualdaten sind nicht nutzbar

- Frage: Kann man anhand der Füllstände auf Fahrten zurückschließen?

2. Daten und Information: Die statistische Trickkiste

- ➔ Die Füllstände der Stationen sind frei verfügbar.
- ➔ Frage: Kann man anhand der Füllstände auf Fahrten zurückschließen?
- ➔ Modell wurde in Coronazeiten für Aufnahme auf Intensivstationen genutzt



2. Daten und Information: Die statistische Trickkiste

Beispiel 2: Fahrrad oder Auto?

- Fahrradnutzung als Kostenersparnis?
- Mitte der 2000 gab es einen rasanten Preisanstieg bei Benzin und Diesel.
- Führen höhere Preise zu Umsteigeeffekten auf öffentliche Verkehrsmittel oder Fahrrad
- Wir untersuchen die Zunahme am Fahrradverkehr in Abhängigkeit vom Benzinpreis
- Es handelt sich un ein natürliches Experiment

Quelle: www.melbourne.vic.gov.au



Bicycle commuting in Melbourne during the 2000s energy crisis:
A semiparametric analysis of intraday volumes

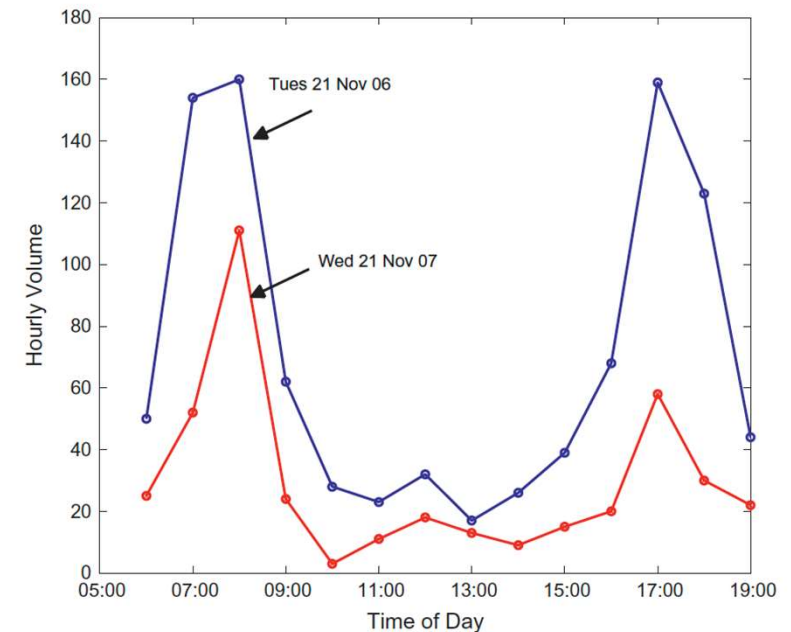
Michael S. Smith^{a,*}, Göran Kauermann^b

^a Melbourne Business School, University of Melbourne, Australia
^b Department of Statistics, Ludwigs-Maximilians-University Munich, Germany

Quelle: Statistika

2. Daten und Information: Die statistische Trickkiste

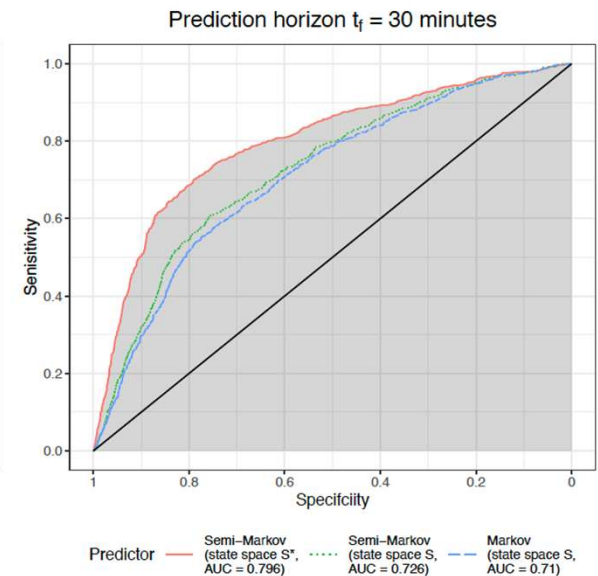
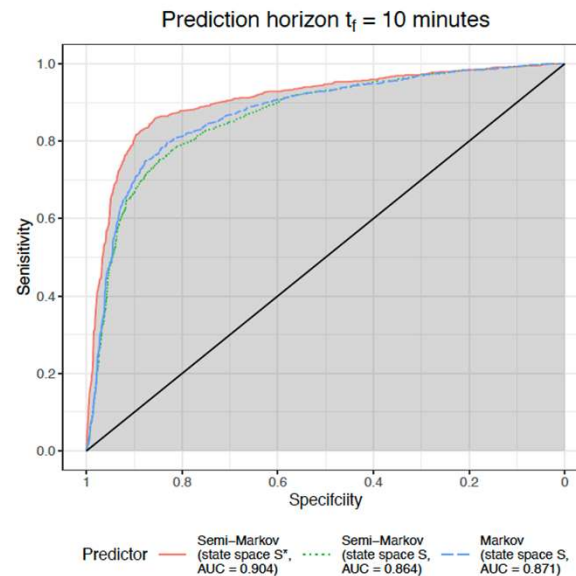
- ➔ Induktionsschleifen zählen vorbeifahrende Fahrräder
- ➔ Die Benutzung des Fahrrads wird vor allem vom Wetter beeinflusst
- ➔ Preiselastizität und Fahrradbenutzung braucht meteorologische Daten
- ➔ Korrigiert man um all weiteren Größen, so war ein **signifikanter Effekt von Benzinpreis** festzustellen
- ➔ Aber, NICHT für alle Induktionsschleifen



2. Daten und Information: Die statistische Trickkiste

Beispiel 3: Smart City Parkplatzsuche

- ➔ Smart Cities: Parkplätze mit Induktionsschleifen
- ➔ Ziel: Navigation zum nächsten Parkplatz
- ➔ Problem: Fluktuation, was ist der beste Parkplatz, den ich anfahren sollte
- ➔ Prognosemodell bei Berücksichtigung der üblichen Verweildauer eines freien Parkplatzes



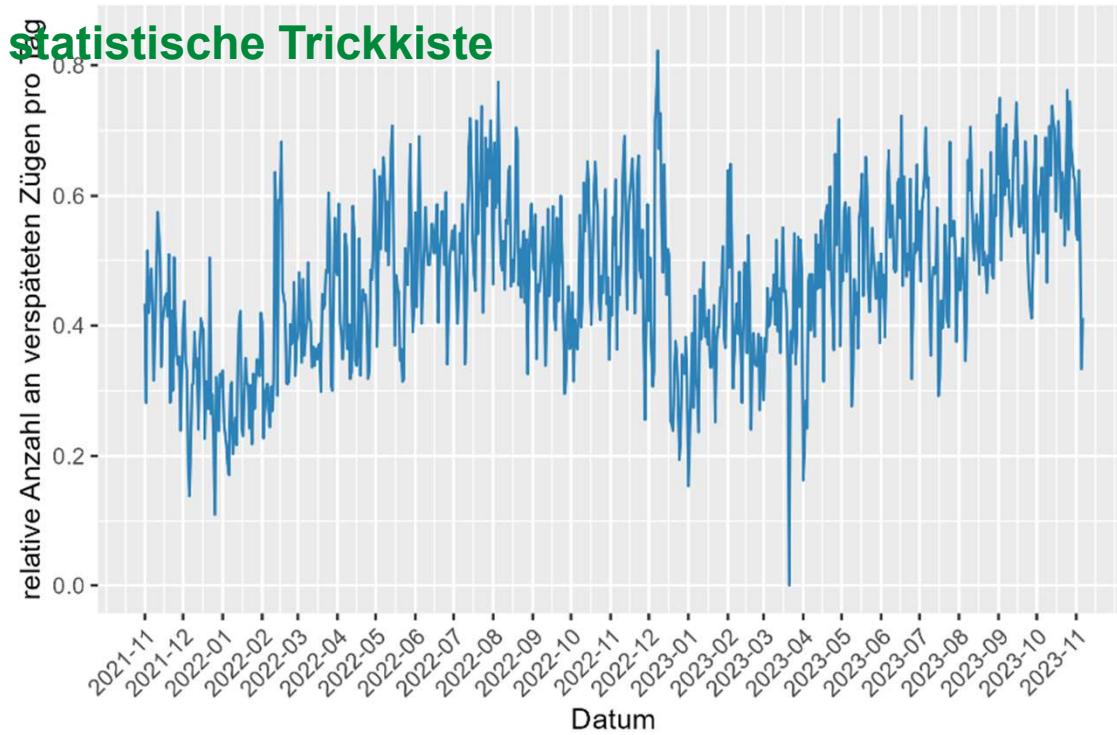
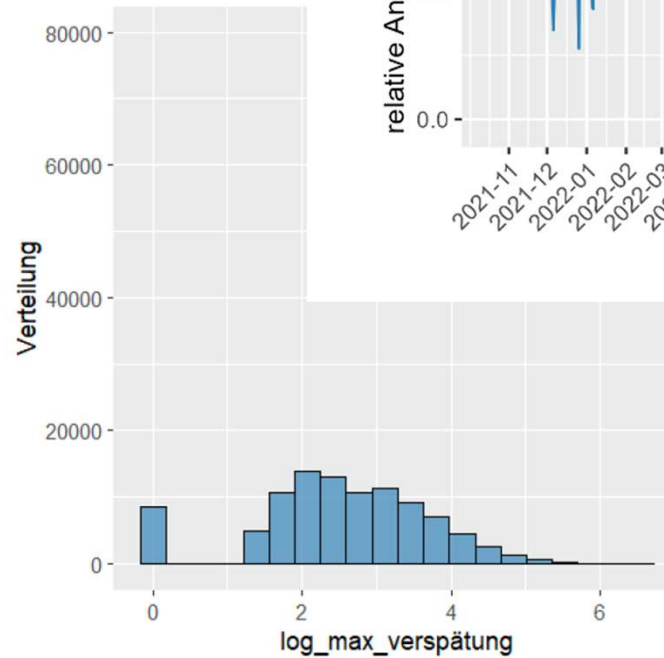
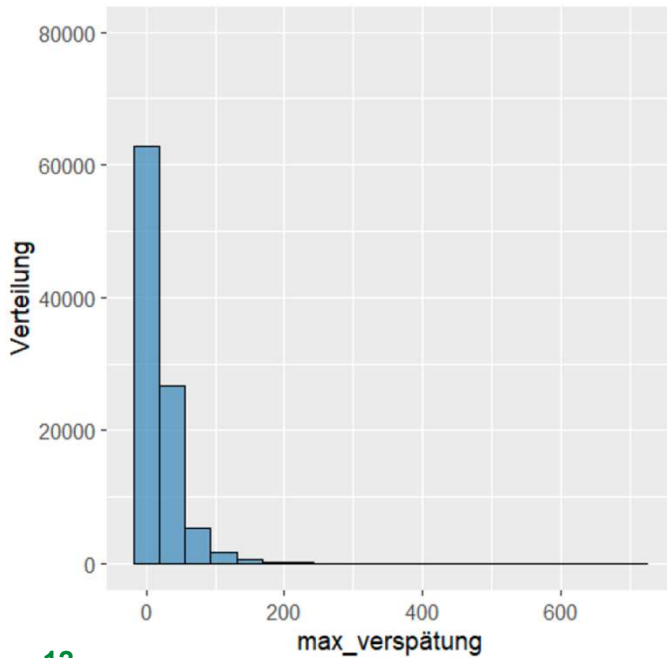
2. Daten und Information: Die statistische Trickkiste

Beispiel 4: Die Bahn

- Bahnverspätungen
- Frage: Gibt es problematische Strecken, hat die Verspätung zugenommen, welche Daten gibt es?
- bahn.de – Fehlanzeige
- www.zugfinder.net
- Die Daten werden NICHT von der Bahn bereitgestellt, sondern aus anderen Quellen „gescrap“
- Man kann die Daten bei zugfinder herunterladen

2. Daten und Information: Die statistische Trickkiste

➔ Bahnverspätungen



Quelle: Bachelorarbeit J.M. Lovric

2. Daten und Information: Die statistische Trickkiste

Beispiel 5: Social Media und Mobilität

- Social Media und Mobilität
- Auswirkungen des Lockdowns Frühjahr 2020 auf die Mobilität und der Einfluss auf die Infektionszahlen
- Daten ????
- Wir benutzen Daten von
Facebook (Data for Good)
- Mobilität = Wahrscheinlichkeit das eine Person aus i eine Person aus j trifft
- Frage: **Reduktion der Infektionen durch Einschränkung der Mobilität?**

ORIGINAL ARTICLE



On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany

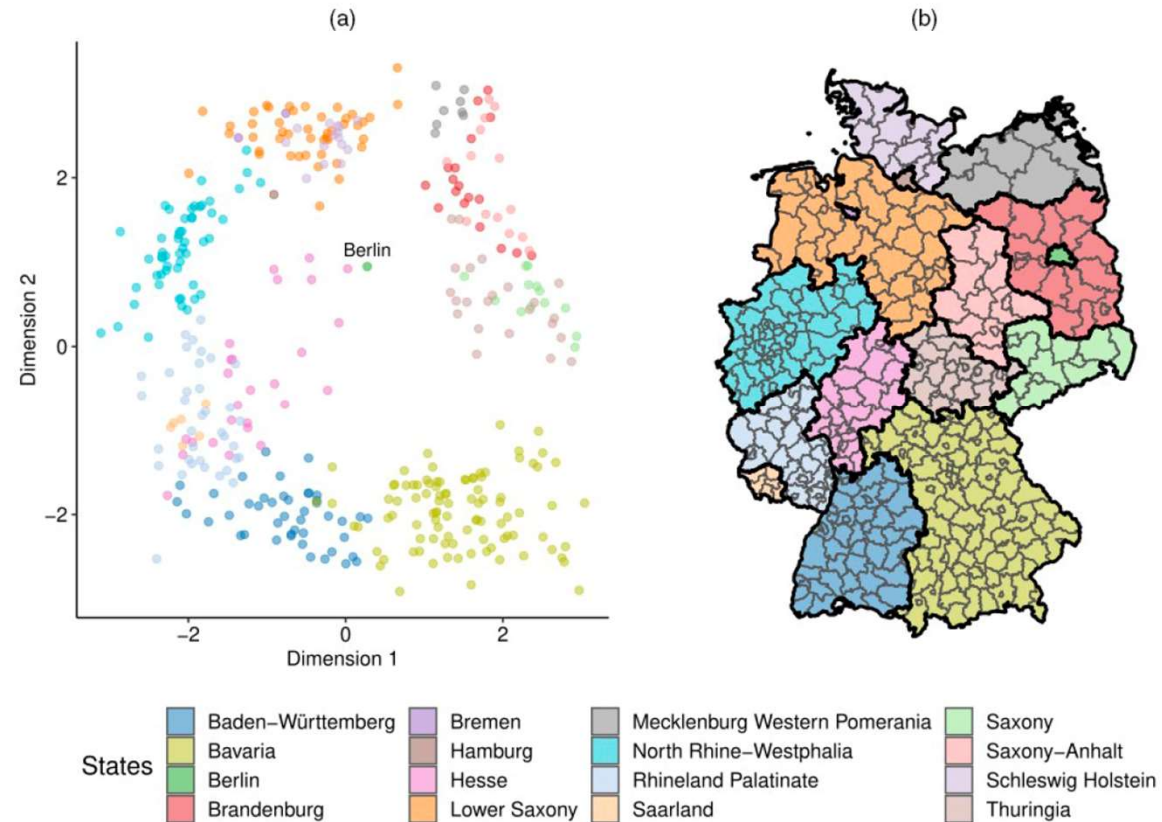
2. Daten und Information: Die statistische Trickkiste

→ Wir sehen signifikante Ergebnisse

- Kein signifikanter Einfluss von Ausgangssperren
- Signifikant negativer Einfluss von lokaler versus weiträumiger sozialer Interaktion.

→ Interessantes „Neben-“Ergebnis: Soziale Nähe in Deutschland

→ Und ???



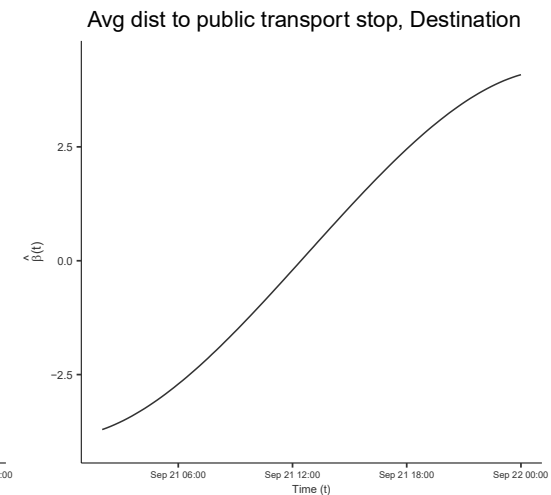
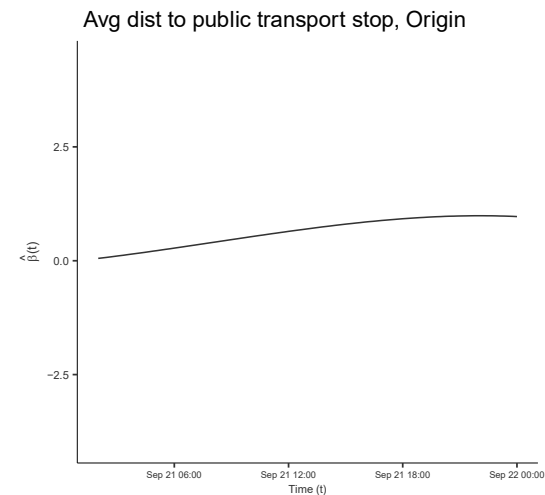
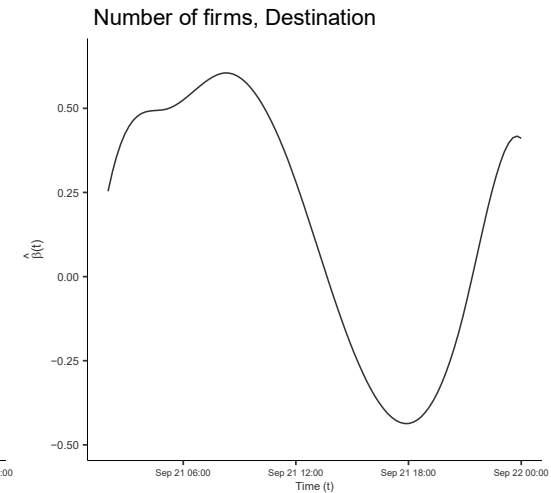
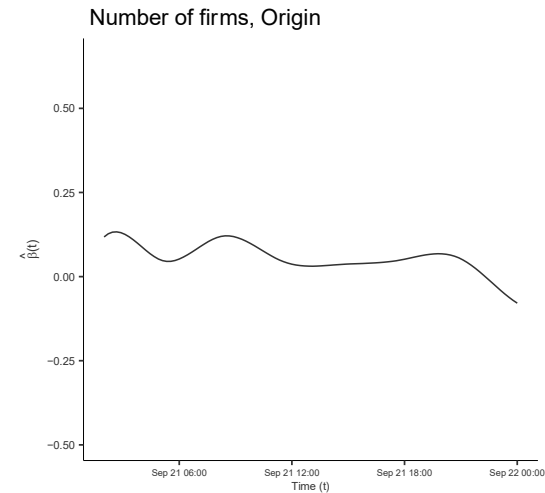
PROF. DR. GORAN RAUERMANN, ILLIAS 2024

2. Daten und Information: Die statistische Trickkiste

Beispiel 6: Mobilität und Telekommunikation

- ➔ Telekommunikationsdate
- ➔ Tagesdaten von Telefonica Kunden
- ➔ Kooperation mit dem ifo Institut
- ➔ Mobilität in München

- ➔ Forschungsfrage: Was beeinflusst Mobilität
- ➔ Gravity Modell mit zusätzlichen Einflussgrößen



3. Daten und Datenschutz in Deutschland, Gegner oder Partner?

→ Was zeigen die Beispiele:

→ Daten

- Melbourne = smart city (seit Jahren)
- Fahrrad Racks = Internationale Provider (keine/wenig deutsche Unternehmen)
- Bahn = smarte Daten Fehlanzeige
- Facebook = data for good
- Telefonica = Forschungsprojekt (1 Tag Daten)

→ Problem:

- Datenschutz

3. Daten und Datenschutz in Deutschland, Gegner oder Partner?

➔ Das Thema Datenschutz aus Sicht eines Experten

„Das Problem ist allerdings nicht das Recht an sich; die DSGVO gilt auch in anderen Ländern. Bei uns ist allerdings deren Interpretation und Anwendung so extrem dezentralisiert („Wimmelbild“) und die Administration wird dominiert von Verwaltungsjuristen. Deren professionelle Ausrichtung ist Risikominimierung ohne Einbeziehung der Nutzenseite.“

„German Angst“

➔ Aber: Das Forschungsdatengesetz und auch ein neues Statistikgesetz stehen auf der Agenda

4. Daten und Qualität: Ein Plädoyer

„I have a dream“

- Freieren Zugang zu Daten für wissenschaftliche Zwecke
- Besserer Zugang zu öffentlichen Daten für kommerzielle Dienste
- Aber: Qualität geht vor Quantität

- Die Verkehrswende braucht Daten und **Datenanalysten**

- Die Novellierung des Forschungsdatengesetzes schreitet voran
- Das Statistikgesetz von 1953 ist nicht mehr zeitgemäß





LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

goeran.kauermann@lmu.de

