M.Sc. in ‚Transportation Systems'

Applied Statistics in Transport Travel Surveys, MID2008

Prof. Regine Gerike
Technische Universität München, mobil.TUM
regine.gerike@tum.de

Munich, 08/11/2011, 15/11/2011

Time Table Updated, 15/11/2011

15.11.2011	8:00-9:30 LE	Real world data input and preparation
15.11.2011	9:45-11:15 L	Studentische Vollversammlung
22.11.2011	8:00-9:30 LE	NO LECTURE
22.11.2011	9:45-11:15 L	NO LECTURE
29.11.2011	8:00-9:30 LE	Descriptive analysis with R, real world data
29.11.2011	9:45-11:15 L	Descriptive statistics
06.12.2011	8:00-9:30 LE	Reserve
06.12.2011	9:45-11:15 L	Distributions
13.12.2011	8:00-9:30 L	Inferential statistics
20.12.2011	8:00-9:30 L	Hypotheses testing
10.01.2012	8:00-9:30 L	Tests, statistical modelling
17.01.2012	8:00-9:30 L	ANOVA
17.01.2012	9:45-11:15 L	Regression
24.01.2012	8:00-9:30 L	Repetition
02.02.2012	Tina Gehlert	Hypothesis-driven data analysis in transport
03.02.2012	Tina Gehlert	Hypothesis-driven data analysis in transport
07.02.2012	10:00-11:00	Exam

Real Data Example MID2008

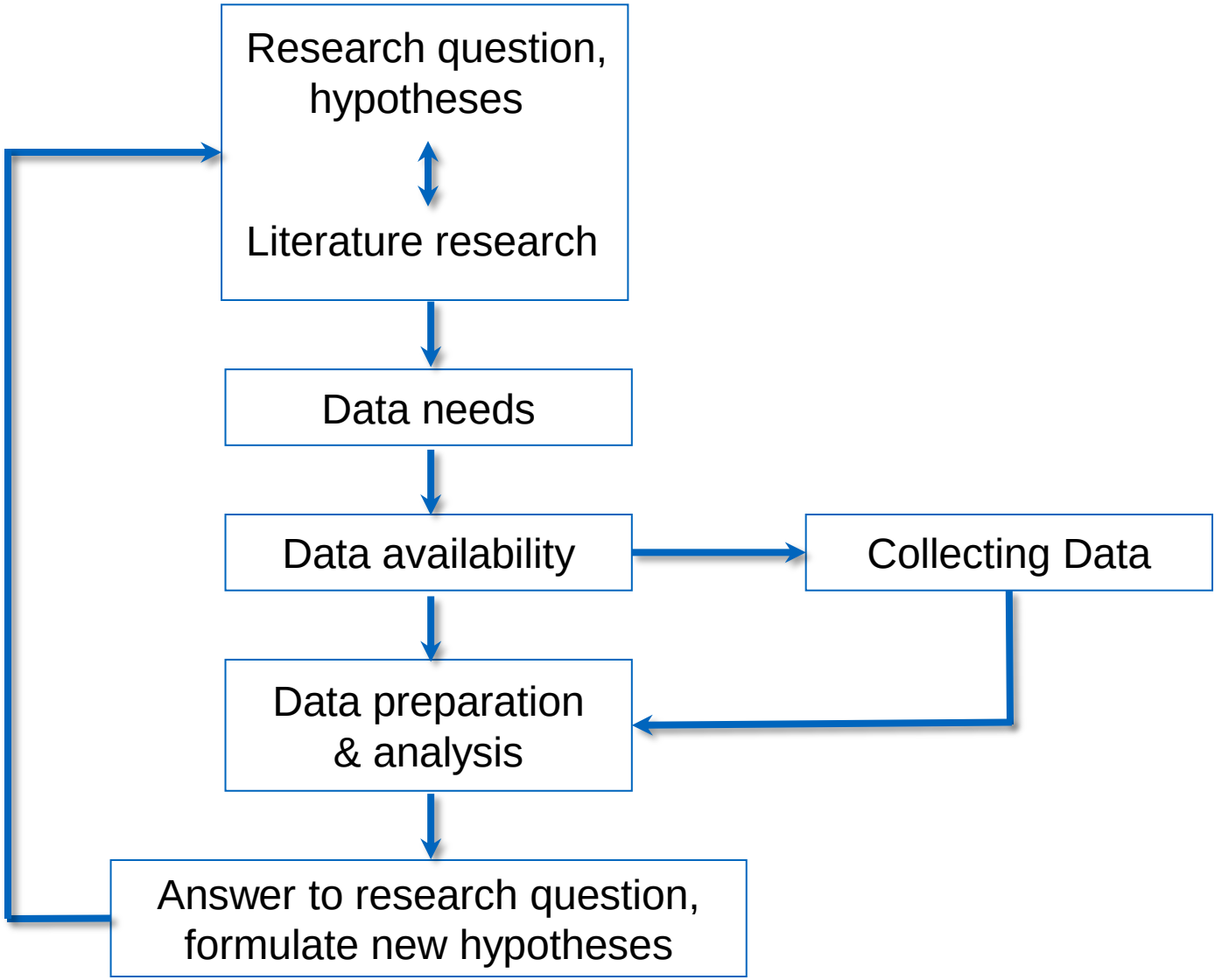
Steps for project:

- Research cycle
- Research questions
- Data Preparation, validation
- Descriptive statistics, visualization
- Tests
- Models, e.g. ANOVA, regression analysis

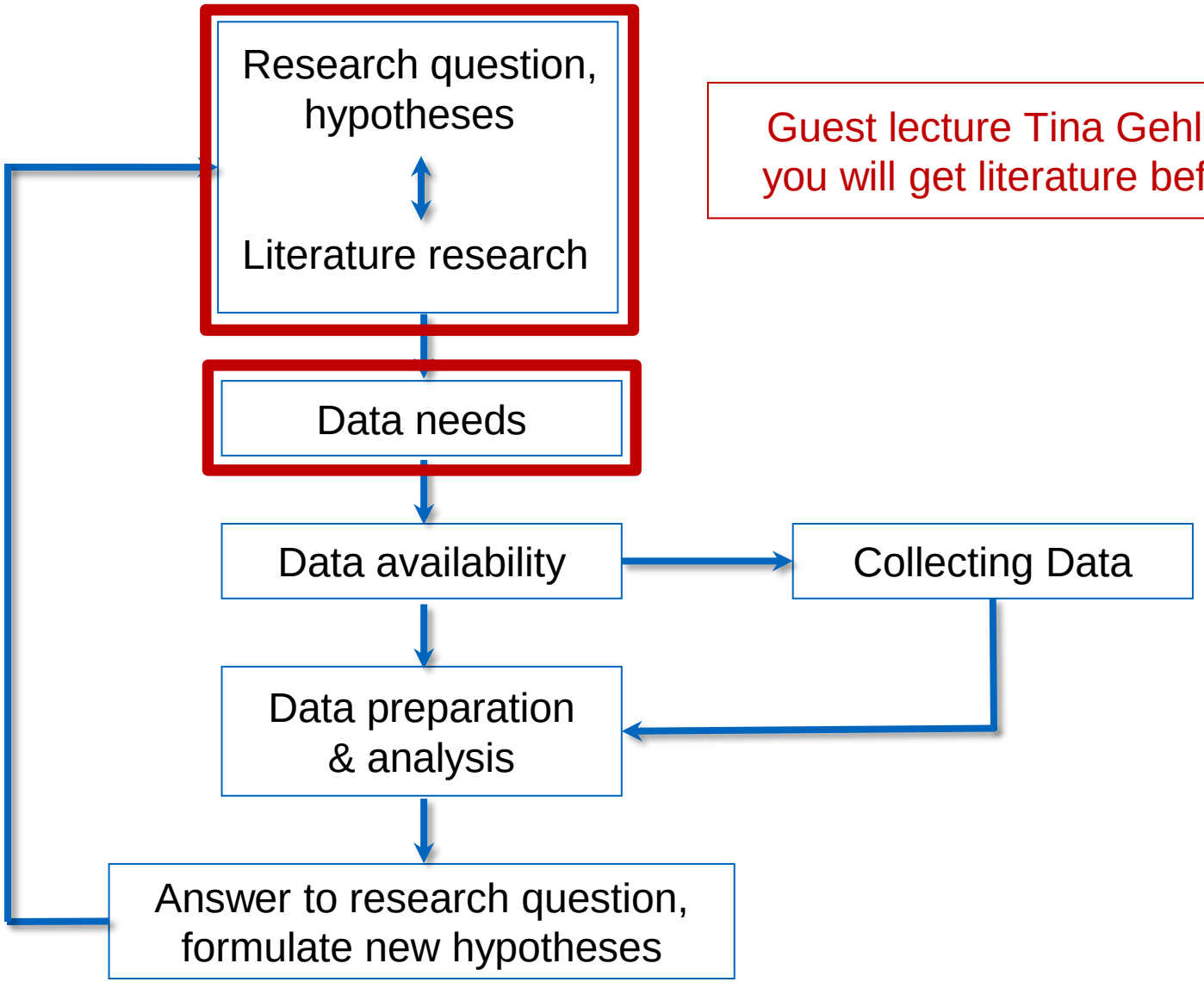
Main Goals:

- Formulate research questions, hypotheses
- Work with real data (read and prepare the data)
- Get familiar with exploratory data analysis (descriptive measures, graphics)
- Get familiar with methods for analyzing your data beyond descriptive statistics

Empirical Research Cycle



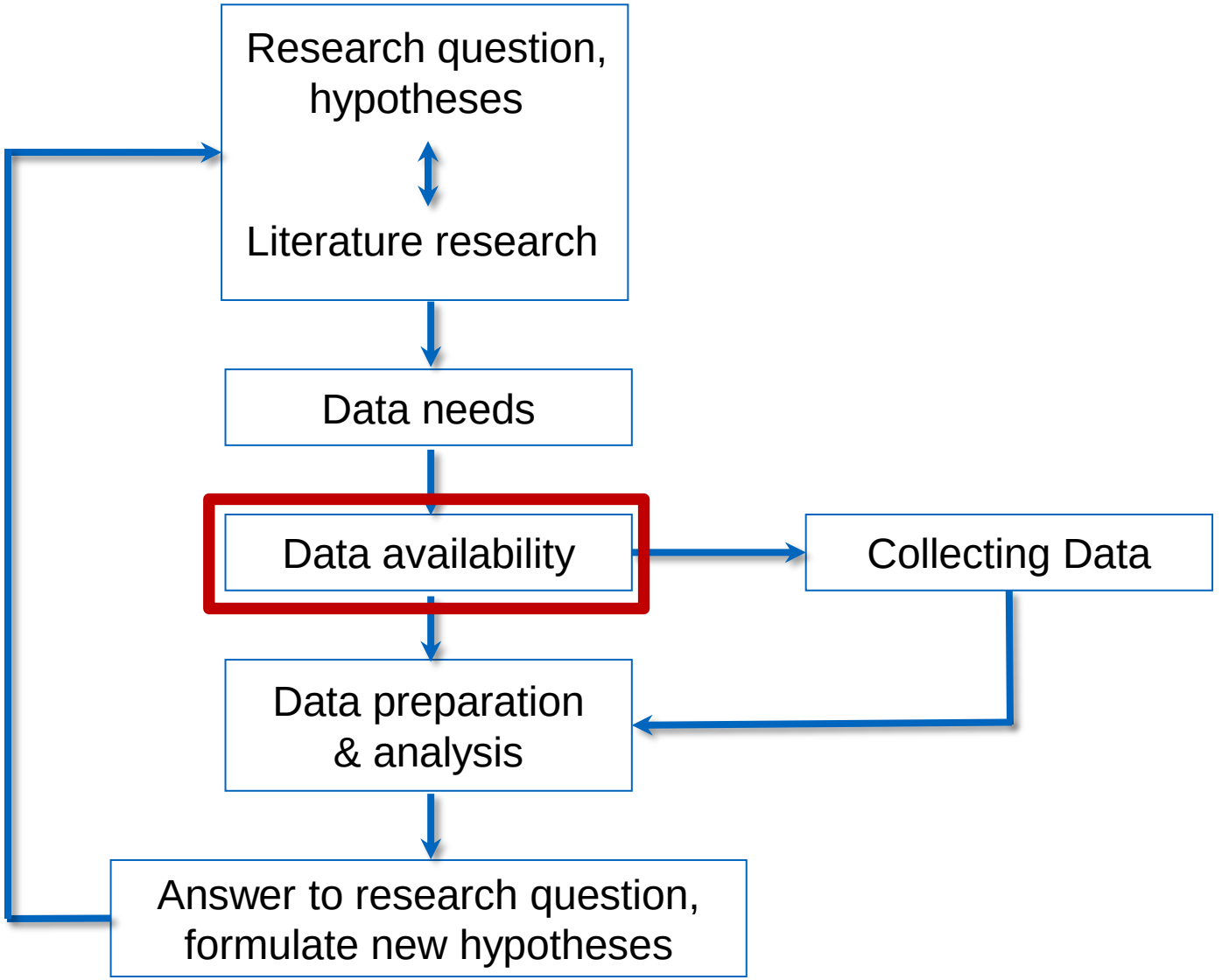
Empirical Research Cycle, Step 1, 2



Step 1: Literature for hypotheses, research questions

- … Guest lecture Tina Gehlert
- Our next three lab sessions: reading and exploring data

Empirical Research Cycle, Step 3



German Travel Surveys: Overview

Two diary day surveys:

- Mobilität in Deutschland: http://www.mobilitaet-in-deutschland.de/
- System repräsentativer Verkehrsbefragungen: http://www.tu-dresden.de/srv/SrV_Web/

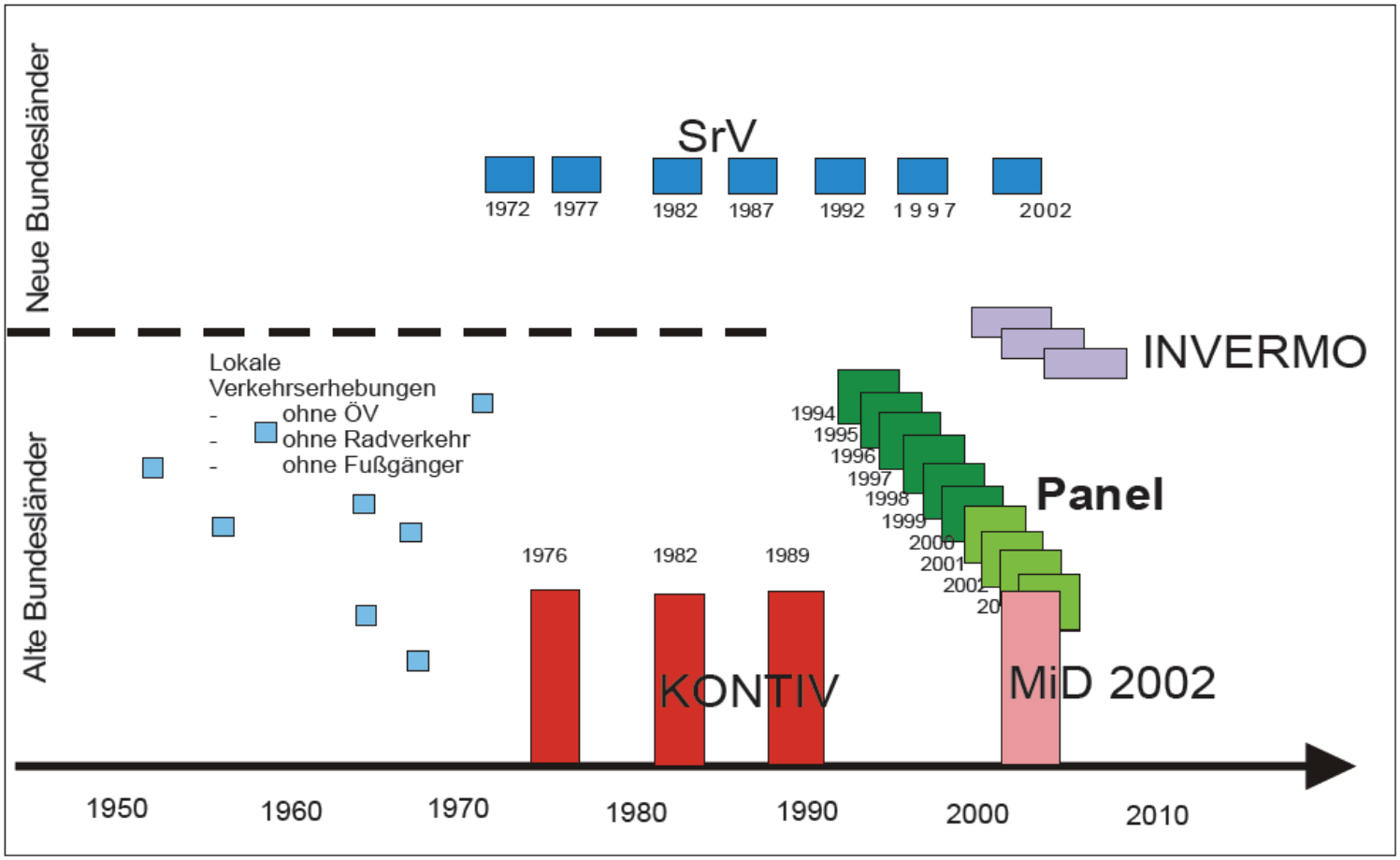
One panel survey:

- Deutsches Mobilitätspanel: http://mobilitaetspanel.ifv.uni-karlsruhe.de/

Data as basis for transport planning and for research

See Kunert-paper for overview of international surveys

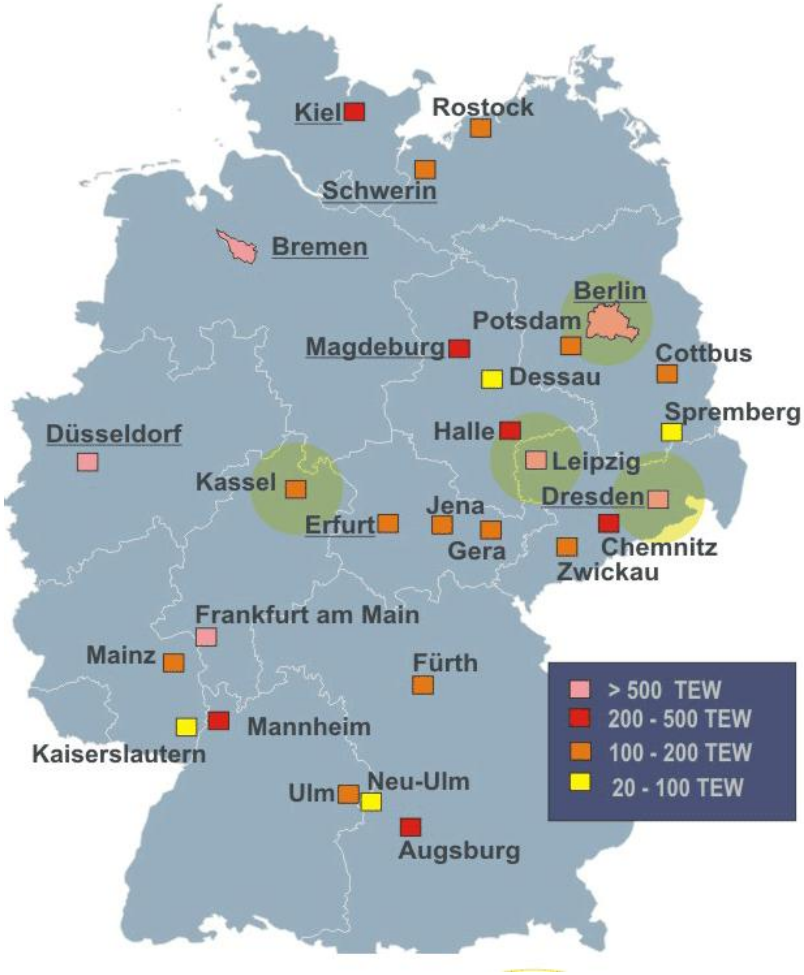
German Travel Surveys: Overview



German Travel Surveys: Mobilität in Deutschland

- Size of random sample net 25,000 households based on registries of residents (covering the whole Germany)
- States or regions add-on by app. another 25,000 households, 100,000 individuals, 300,000 trips
- Collection of information of the whole household
- Survey guided by fixed diary-dates and lasting 12 months (weekdays and weekend-days)
- Non-response-study
- Written questionaires, telephone interviews, online questionaires
- Years 2002, 2008
- Similar surveys: 1976, 1982, 1989 "KONTIV" (Kontinuierliche Erhebung zum Verkehrsverhalten)

System repräsentativer Verkehrsbefragungen

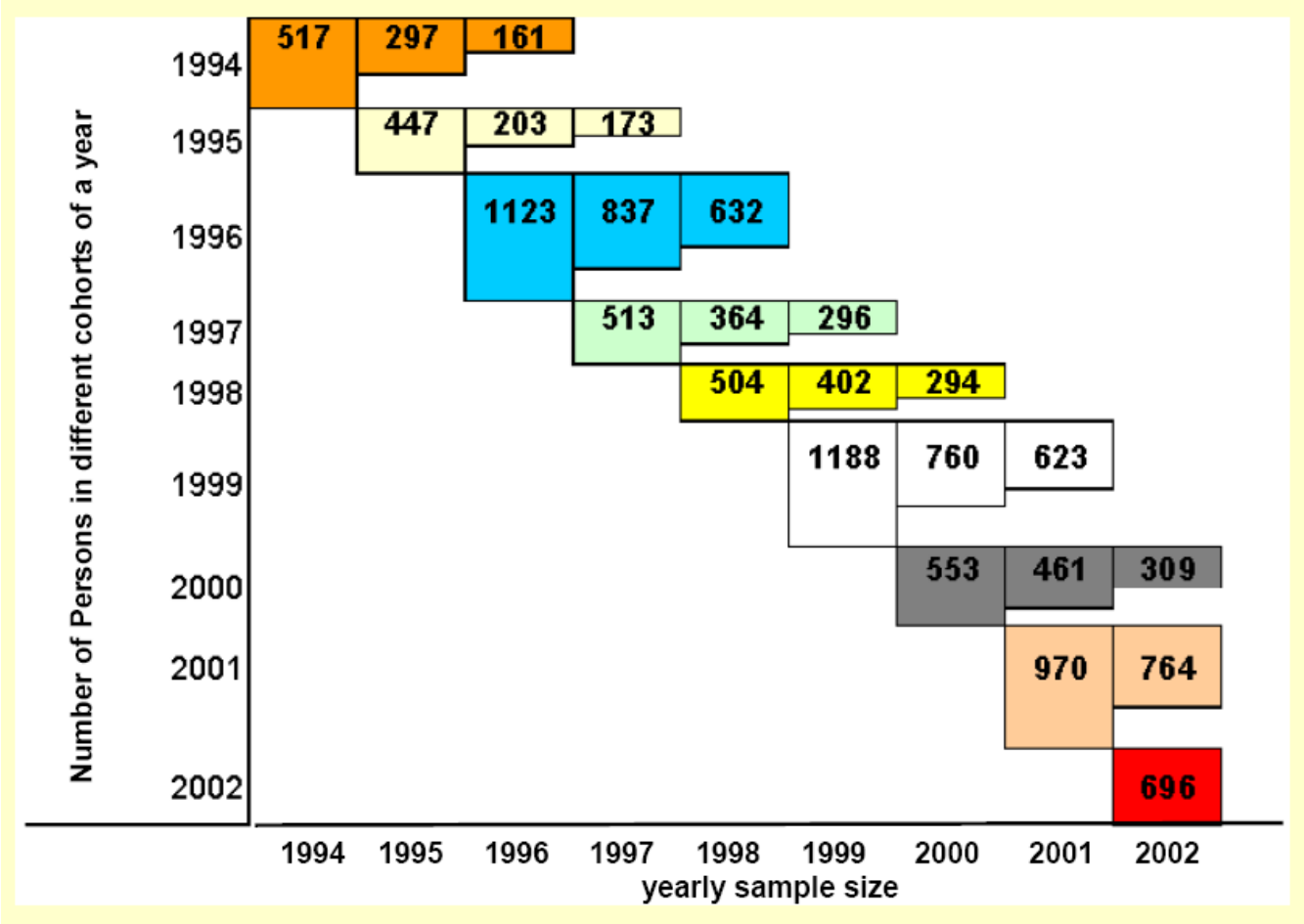


Participating cities in 2008,
http://www.tu-dresden.de/srv/SrV_Web/

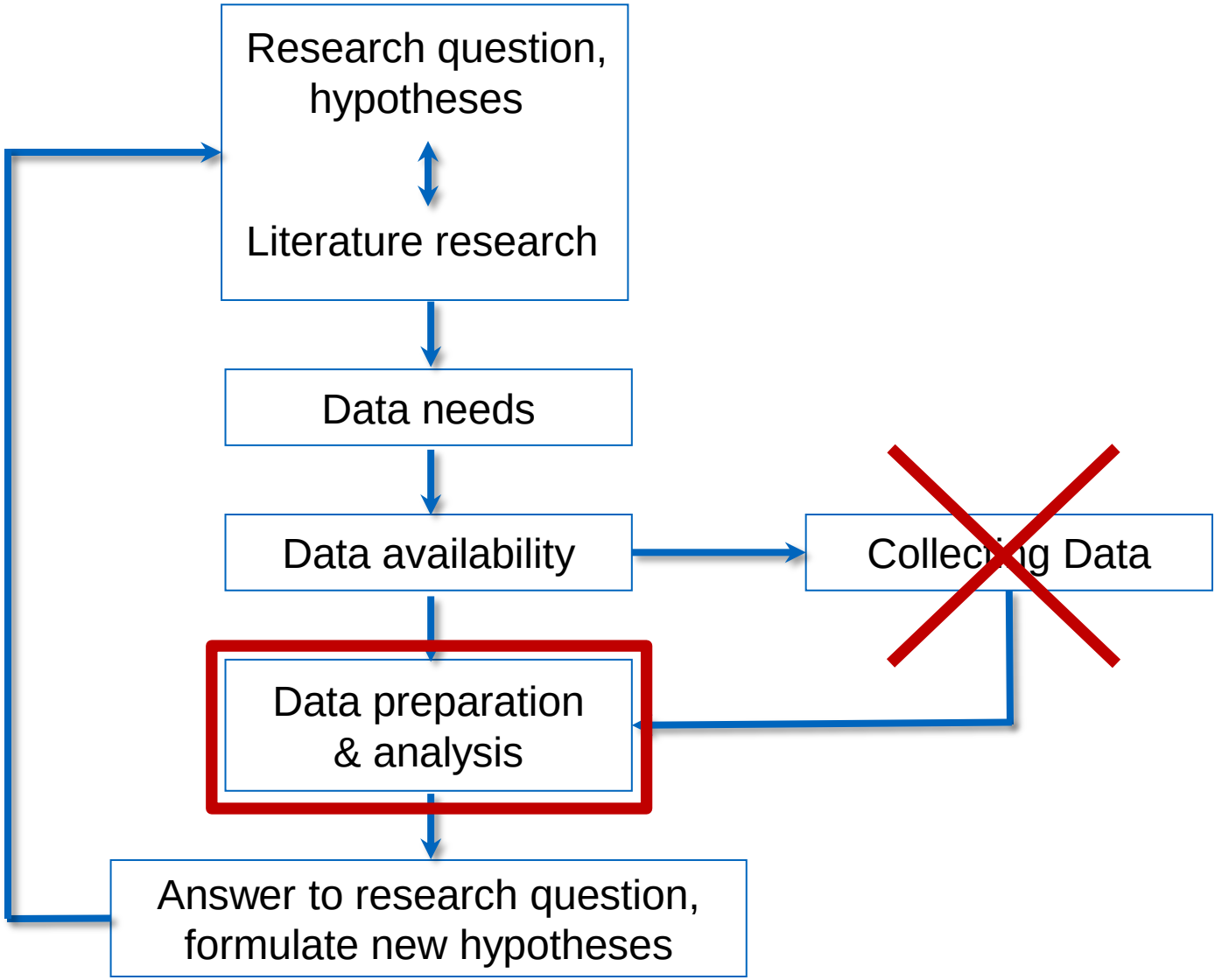
German Mobility Panel, Deutsches Mobilitätspanel

- Mobility behaviour of one complete week (longitudinal continuous approach) - diary for one week
- Yearly repetition (one week in autumn)
- Mobility of complete households
- Rotating sample (3 years of participation)
- No geocoding of destinations

German Mobility Panel



Empirical Research Cycle, Step 4

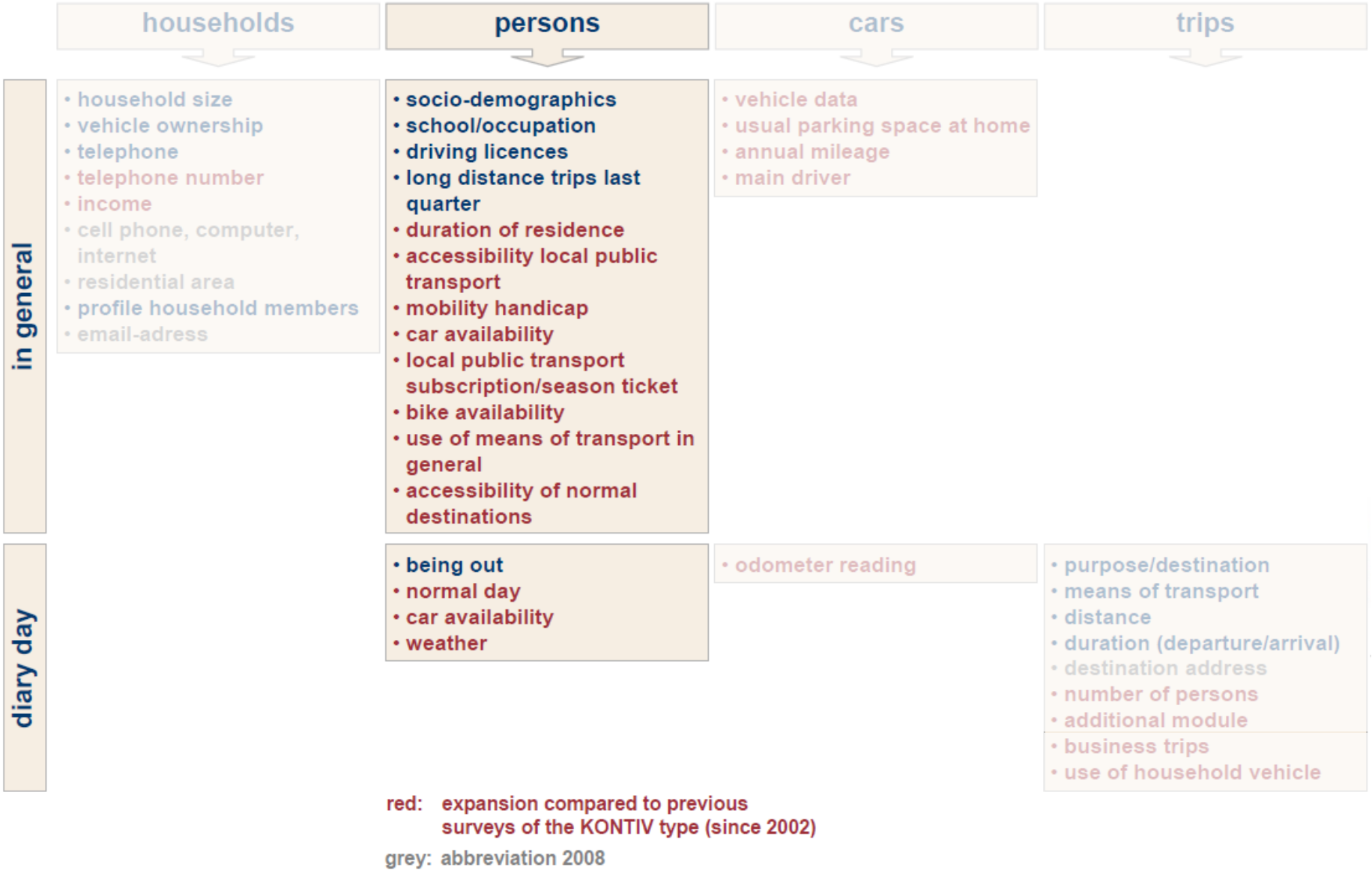


Decision on Data Set: Mobilität in Deutschland 2008

- One diary day
- Weekdays and weekend days during a whole year
- All region types
- Covering the whole household (including children, complete households in 81% of the cases)

- http://www.mobilitaet-in-deutschland.de/engl%202008/

MiD2008: Data sets, Structure



MiD2008: Material for your work, you get:

Textfiles, the data:

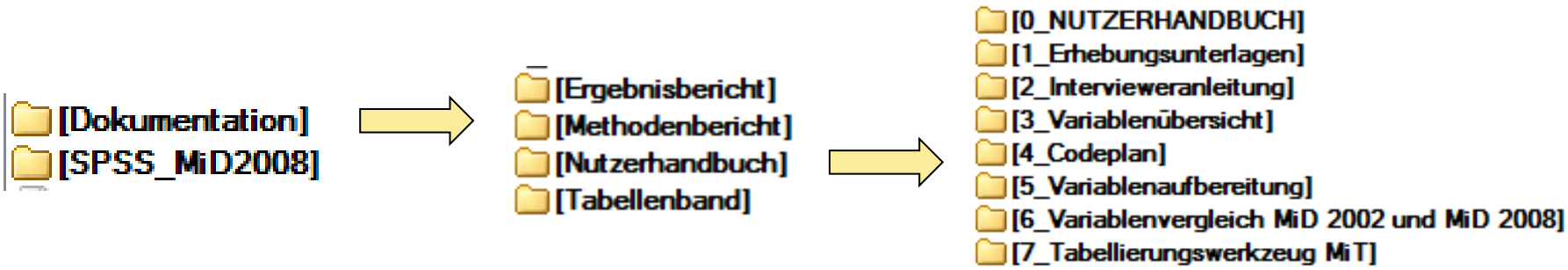
- MiD2008_PUF_Personen.dat, one case/observation = one person
- p2008extendedAS.Rdata (= MiD2008_PUF_Personen.dat but with more variables)

Codeplan:

- MiD2008EnglCodeplan20111114extendedAS.xlsx

Additional material:

- Information on important variables, 2011-11-10_MiD2008_Information_en.docx
- R-code for today: MID2008AppliedStatistics_20111114.R



Literature on MID2008, Travel Surveys

- DLR, Infas (2007) Mobilität in Deutschland 2008 – MID 2008 (Mobility in Germany), Information Meeting Add-Ons, Bonn, 24 July 2007, see "Mobility in Germany 2008_Projects Presentation.pdf“, overview of the MID2008 approach
- Kunert, U.; Kloas, J.; Kuhfeld, H. (2001) Design Characteristics of National Travel Surveys, An International Comparison for ten Countries, see Kunert2001.pdf: Overview of KONTIV (former version of MID) and other travel surveys
- Questionnaire for trips, see 2011-11-10_trip-sheet.xlsx

- United Nations (2005) Household Sample Surveys in Developing and Transition Countries, http://unstats.un.org/unsd/hhsurveys/, see UnitedNations2005_Household_surveys.pdf, 655 pages, not relevant for the exam, for all who are interested in surveying

Example MID2008, steps

3 sessions in the computer lab:

1. Reading the data, recoding, plausibility checks, filter checks
2. Descriptive statistics
3. Open

Guest lecture Tina Gehlert

Exploring the data – steps for data preparation

- Check individual variables for extreme and implausible values
- Check cross variable consistency (e.g. compute speed from trip distance and trip time, check with transport mode)
- Compute new variables that are needed for the analysis

- Most important:
- **Never work with original data! Use copies for your analyses!**

Thank you for your attention.

Regine Gerike
Technische Universität München
mobil.TUM
Office: 1753
Tel +49.89.289.28575
regine.gerike@tum.de

MiD2008

- Relevant variables:
- Number of trips on the diary day: wege1
- Travel time diary day: anzmin
- Travel distances diary day: anzkm
- Yearly distances travelled by car [km/year]: fahrlj_h
- CO_2-Emissions: co2tag_h (Household, diary day), co2tag_p (Person, diary day), co2weg (per trip)
- Household income: hheink

- Household variables: household type and household size, number of employed household members, number of driving licences, number of cars/bikes/motor cycles, reasons for not owning a car, type of region
- Person variables: car/bike availability, frequency usage car/bike/PT/airplane, gender, age, distances to and accessibility of facilities, employment status, education, type of PT-tickets, driving licence, availability handies/computers, mobility restrictions, type of region