Daytime Population estimations based on Mobile Phone Metadata

Martijn Tennekes and May Offermans

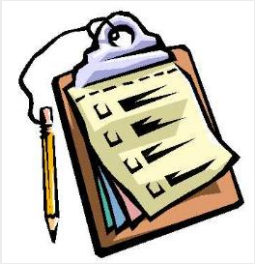
Statistics Netherlands

Big Data

from an Official Statistics point of view

Three types of data:

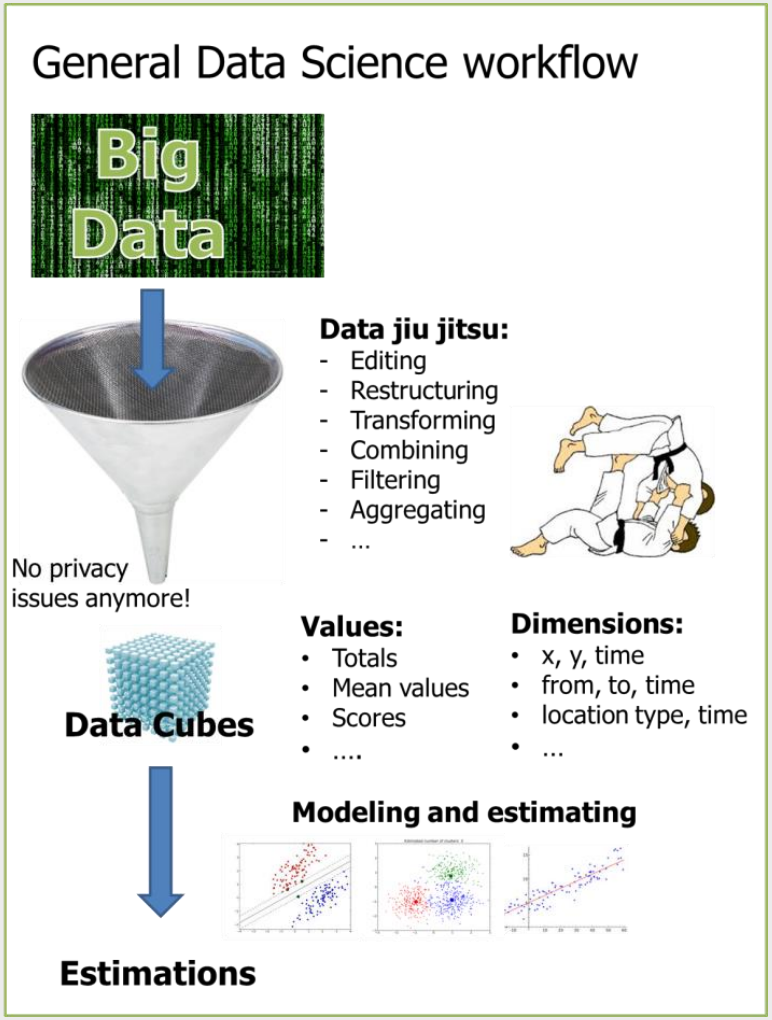
1. Survey data = data collected by SN with questionnaires



2. Admin data = administrative (register) data collected by third parties such as the Tax Office

3. Big data = machine generated data of events

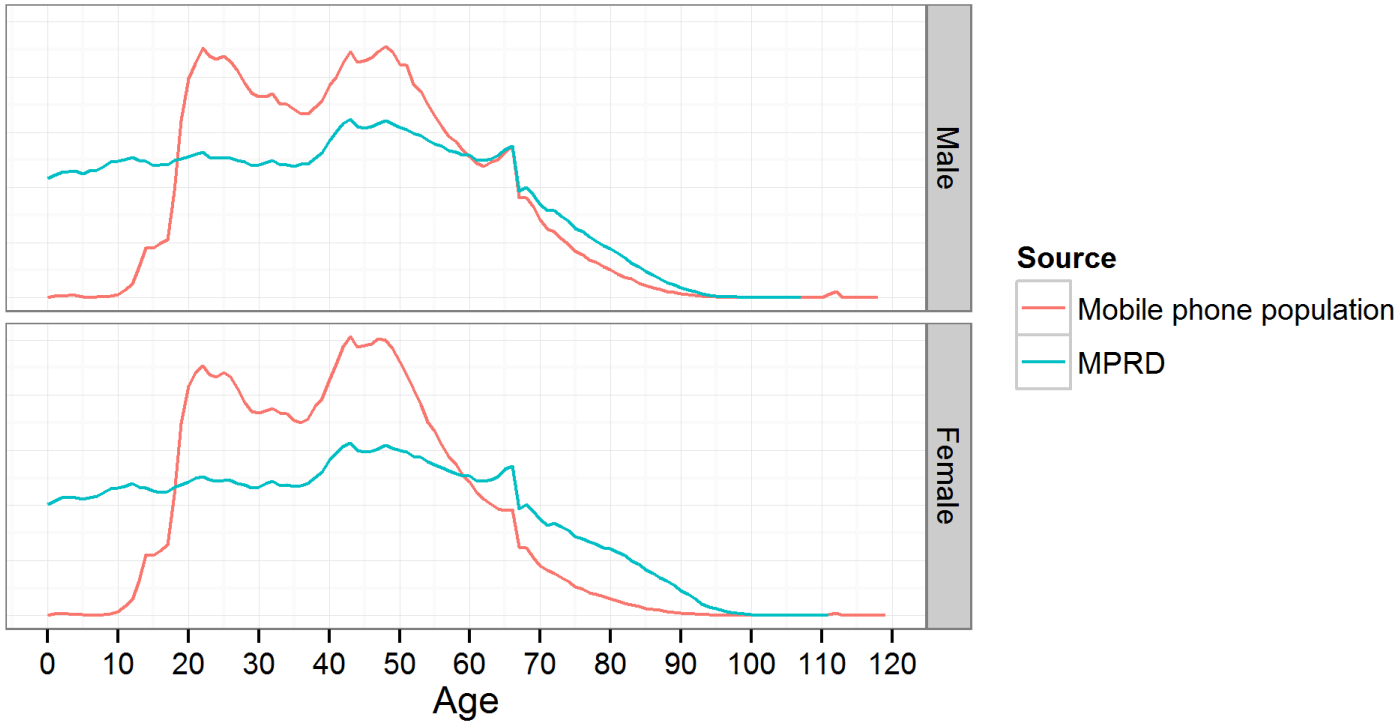
Big data approach



Mobile phone metadata

- Pilot study with Vodafone, a provider with market share of 1/3 in the Netherlands.
- Aggregated data is queried by intermediate company Mezuro and delivered to SN. Privacy is guaranteed!
- Applications: daytime population, tourism statistics, economic activity, mobility studies, etcetera.

Mobile phone population



MPRD (Municipal Personal Records Database) = Dutch population

Subpopulations model

Mobile phone metadata weighted to the MPRD.

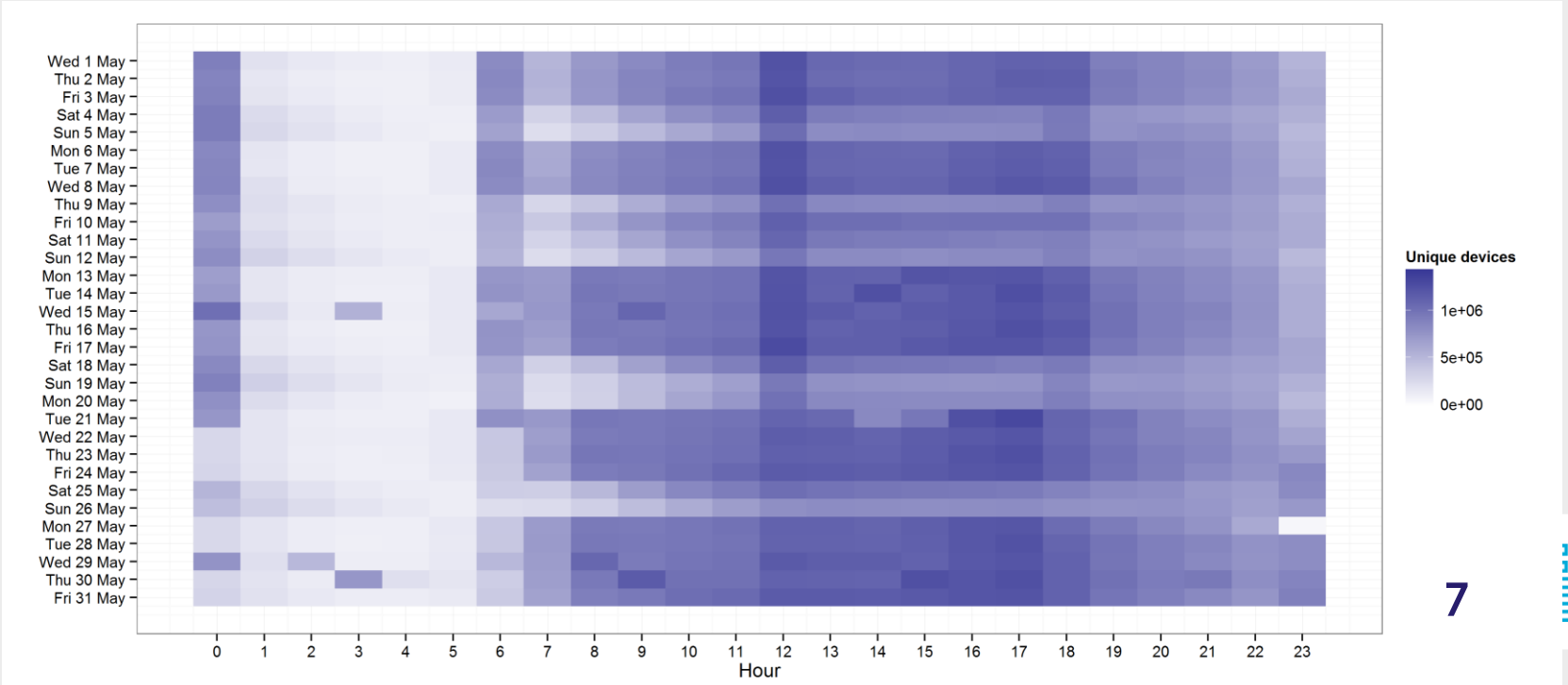
MPRD data & Education Registers.

MPRD data only.

Mobile phone metadata

Event Datail Records (EDR) contain metadata on mobile phone events (i.e. call, SMS or data transfer).

Aggregated table: number of unique devices X time period X current region X residential region.



Weighting method

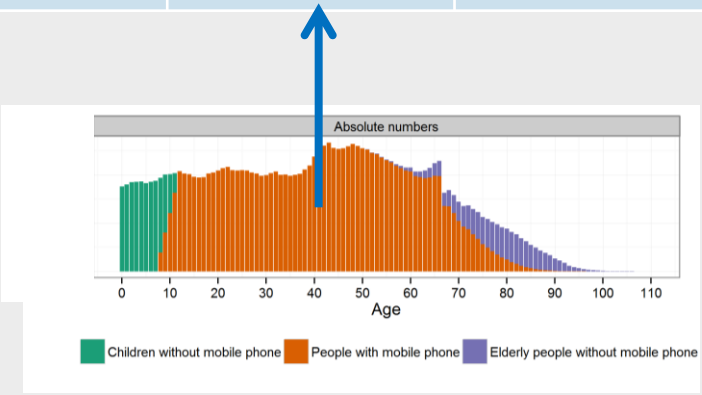
Example: suppose there are only 3 regions in the Netherlands: Amsterdam, Boskoop and Castricum

	Residence				
Current region at time *t*		Amsterdam	Boskoop	Castricum	
	Amsterdam	199,000	1,000	4,000	
	Boskoop	500	3,500	0	
	Castricum	500	500	16,000	

Weighting method (2)

Example: suppose there are only 3 regions in the Netherlands: Amsterdam, Boskoop and Castricum

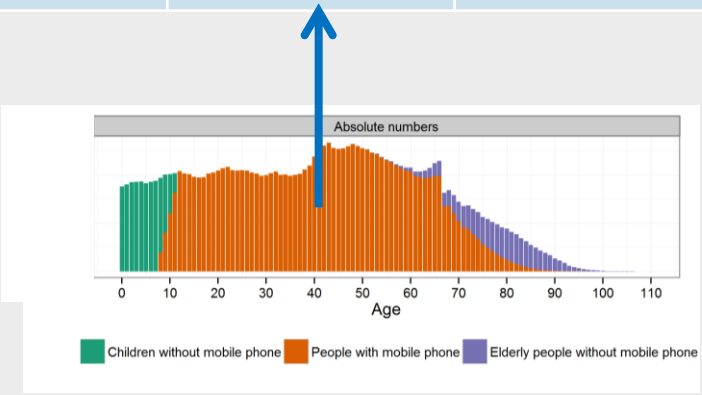
	Residence				
Current region at time *t*		Amsterdam	Boskoop	Castricum	
	Amsterdam	199,000	1,000	4,000	
	Boskoop	500	3,500	0	
	Castricum	500	500	16,000	
	MPRD total	**600,000**	**15,000**	**30,000**	



Weighting method (3)

Example: suppose there are only 3 regions in the Netherlands: Amsterdam, Boskoop and Castricum

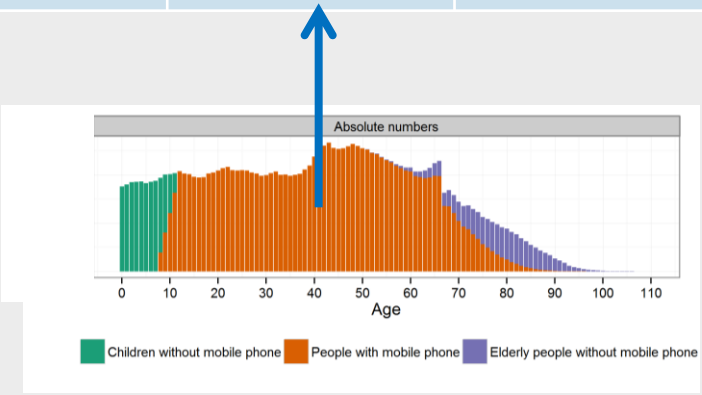
	Residence				
Current region at time *t*		Amsterdam	Boskoop	Castricum	
	Amsterdam	596,000	3,000	6,000	
	Boskoop	2000	10,500	0	
	Castricum	2000	1,500	24,000	
	MPRD total	**600,000**	**15,000**	**30,000**	



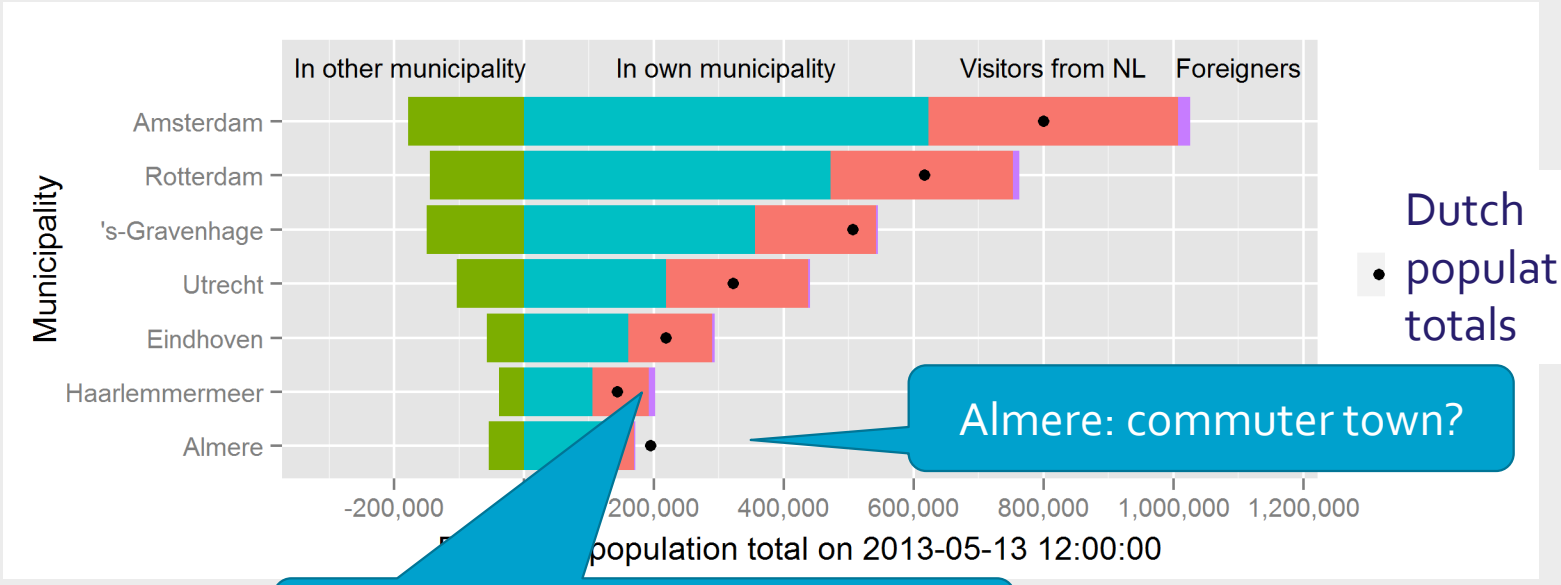
Weighting method (4)

Example: suppose there are only 3 regions in the Netherlands: Amsterdam, Boskoop and Castricum

	Residence				
Current region at time *t*		Amsterdam	Boskoop	Castricum	**DTP total**
	Amsterdam	596,000	3,000	6,000	605,000
	Boskoop	2000	10,500	0	12,500
	Castricum	2000	1,500	24,000	27,500
	MPRD total	**600,000**	**15,000**	**30,000**	



Daytime population results

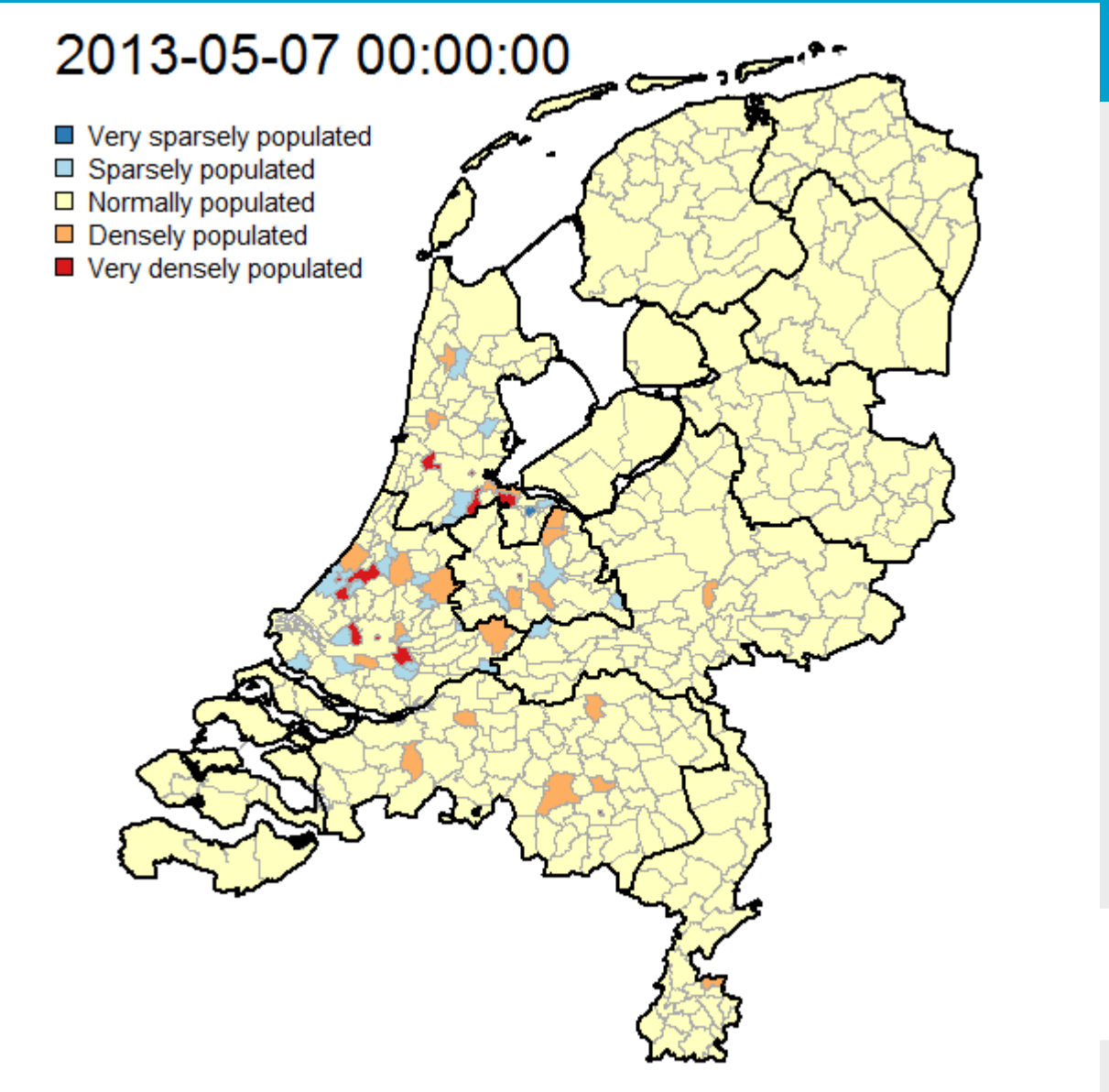


Dutch population totals

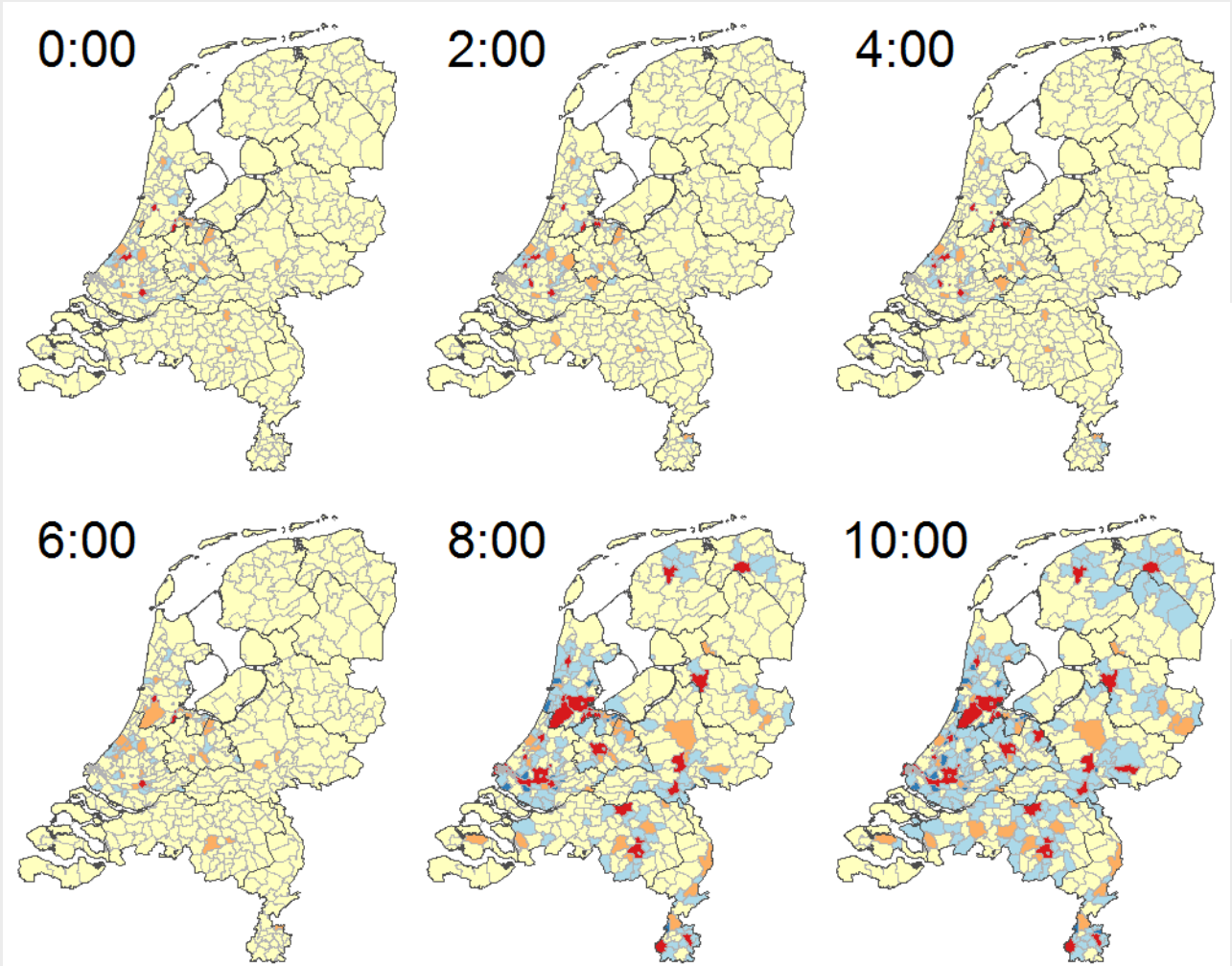
Almere: commuter town?

Foreigners at Schiphol Airport

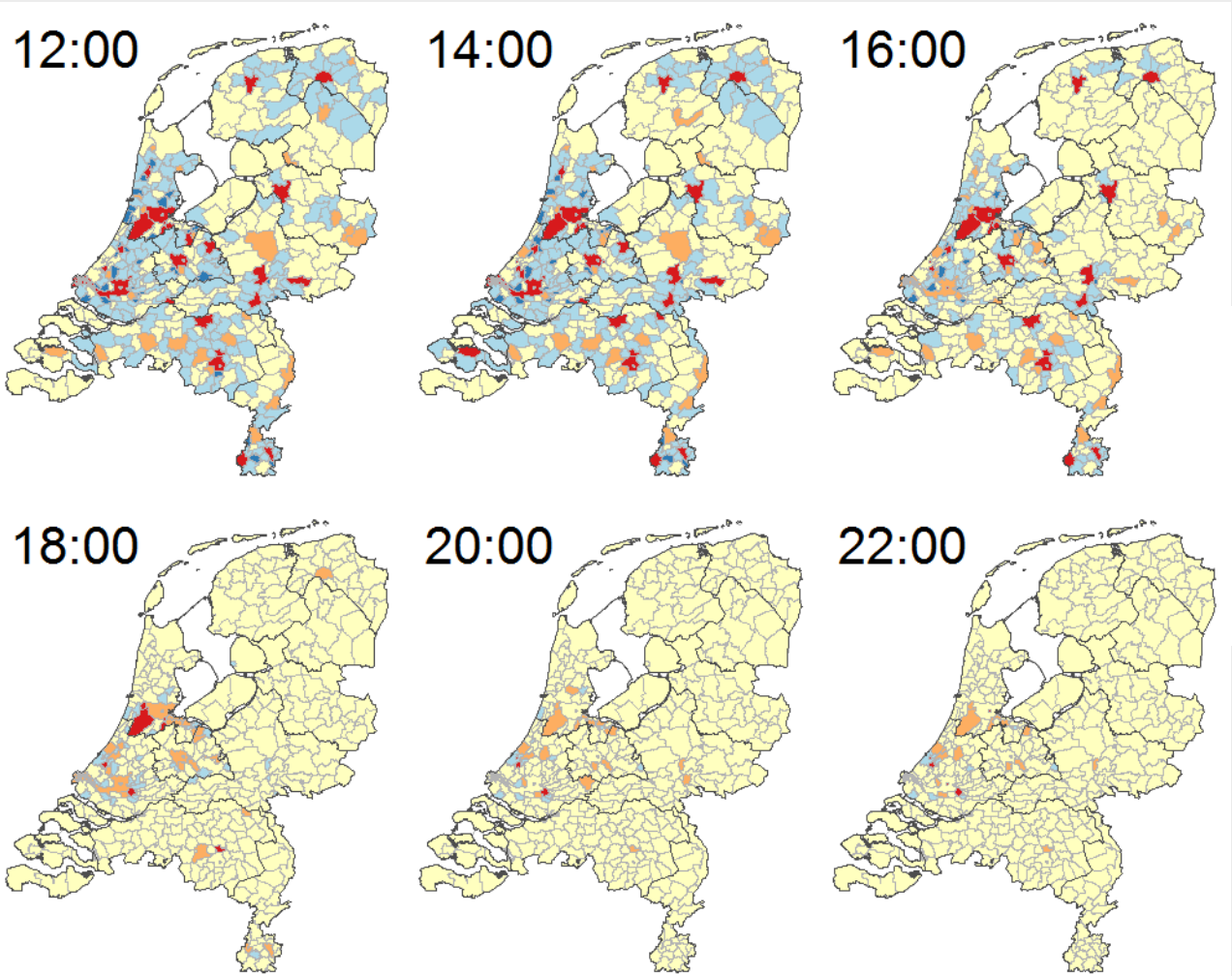
Day time population (relative)



Day time population (relative)

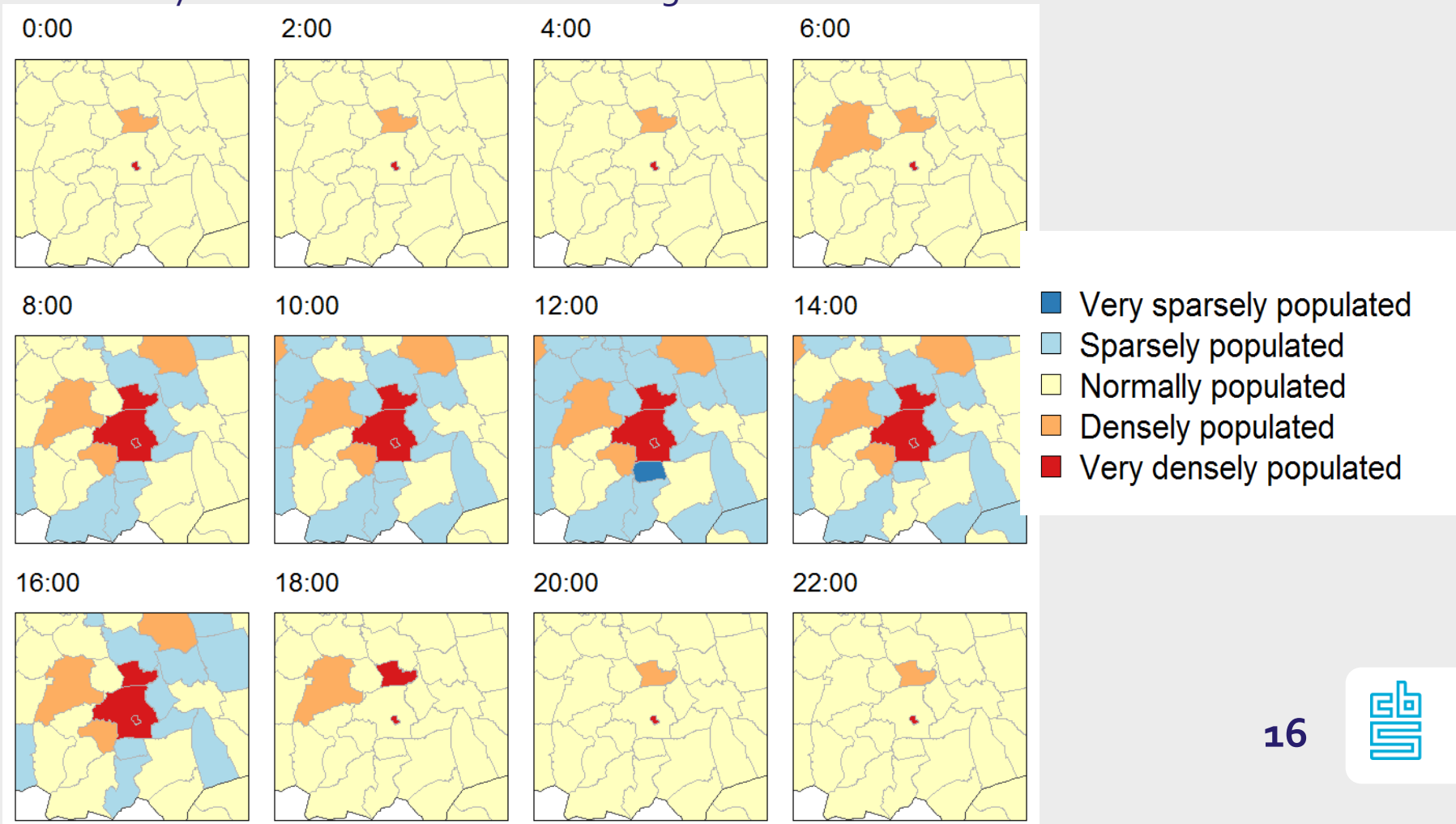


Day time population (relative)



Day time population (relative)

City of Eindhoven and surrounding towns



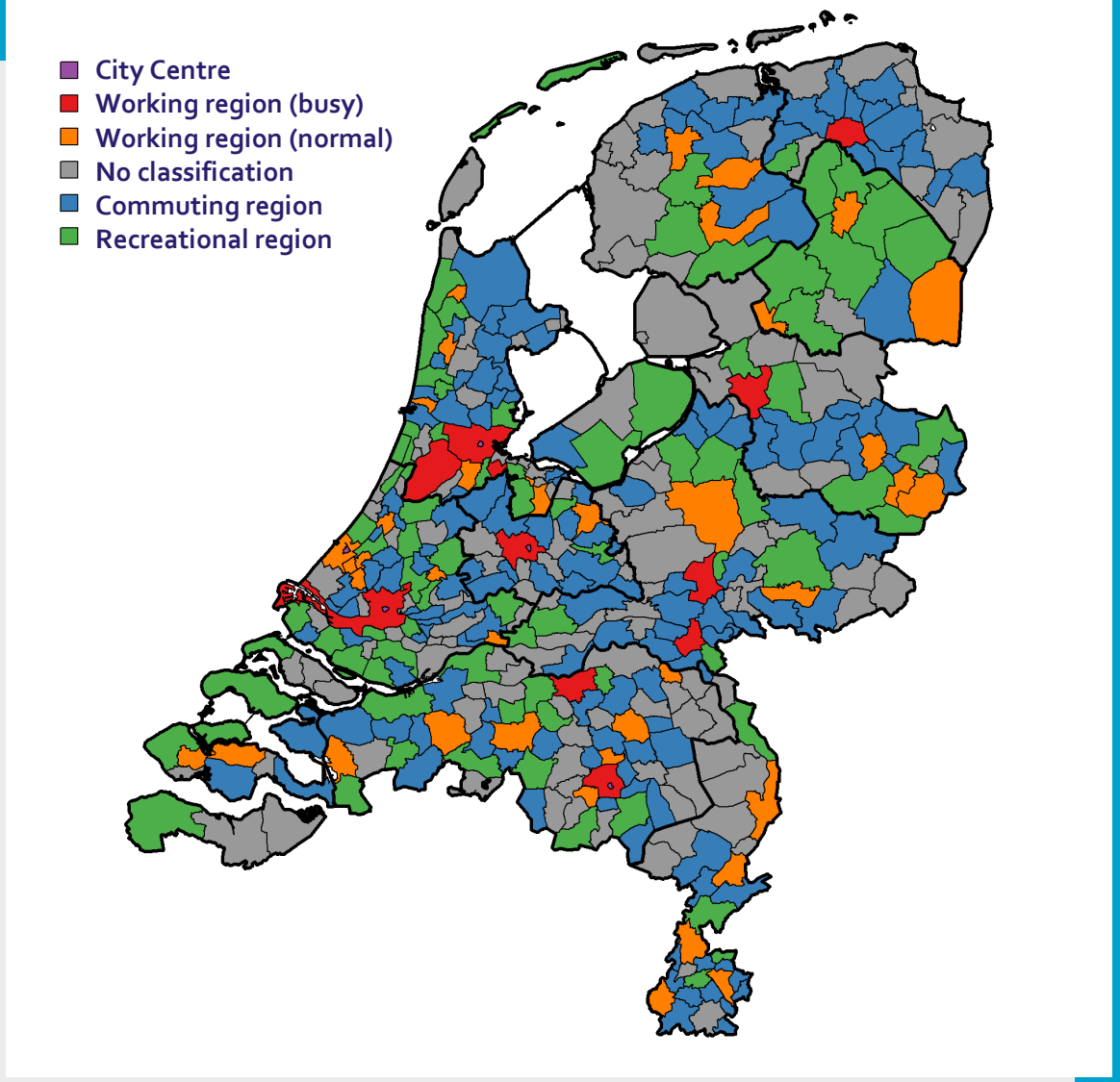
Day time population – Region profile

K-means clustering

Work = daytime vs. night-time during working weeks

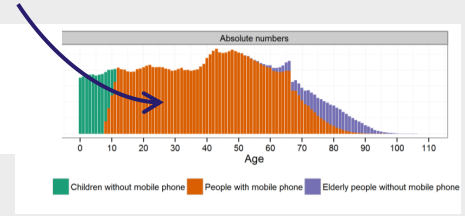
Weekend = weekends activity

Holiday = May holiday activity



Further Research

- Assumptions for the middle part of the population:
  - Number of devices per person = 1
  - All devices are used frequently



- Validation of subpopulation model
- WIFI offload
- Improvement on roaming estimations:
  - Foreigners in the Netherlands
  - Dutch people abroad

Questions?