

# Traffic loop data for transport statistics

*Peter Struijs*

International Conference on Big Data for Official  
Statistics, Beijing, China, 28 – 30 October 2014



Statistics  
Netherlands

# Topics covered

- Characteristics of the data source
- Issues when using traffic loop data
- Solutions to issues
- Results

# The main roads

Highways per NUTS3



# Road sensors

## Road sensor (traffic loop) data

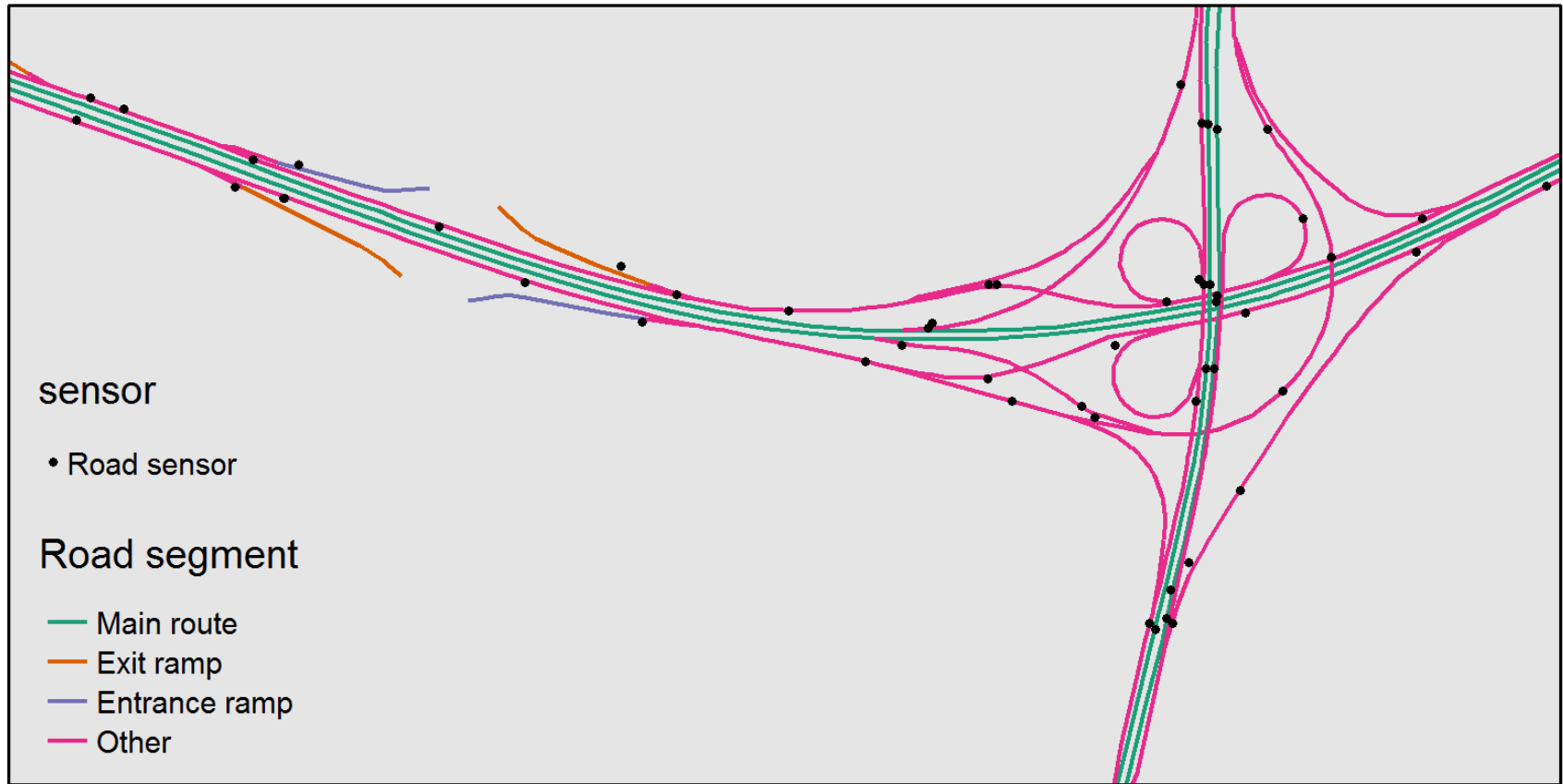
- Each minute (24/7) the number of passing vehicles is counted in around 20.000 highway 'loops' for different length classes
- No identification of vehicles
- Big Data: around 230 million records a day



Locations



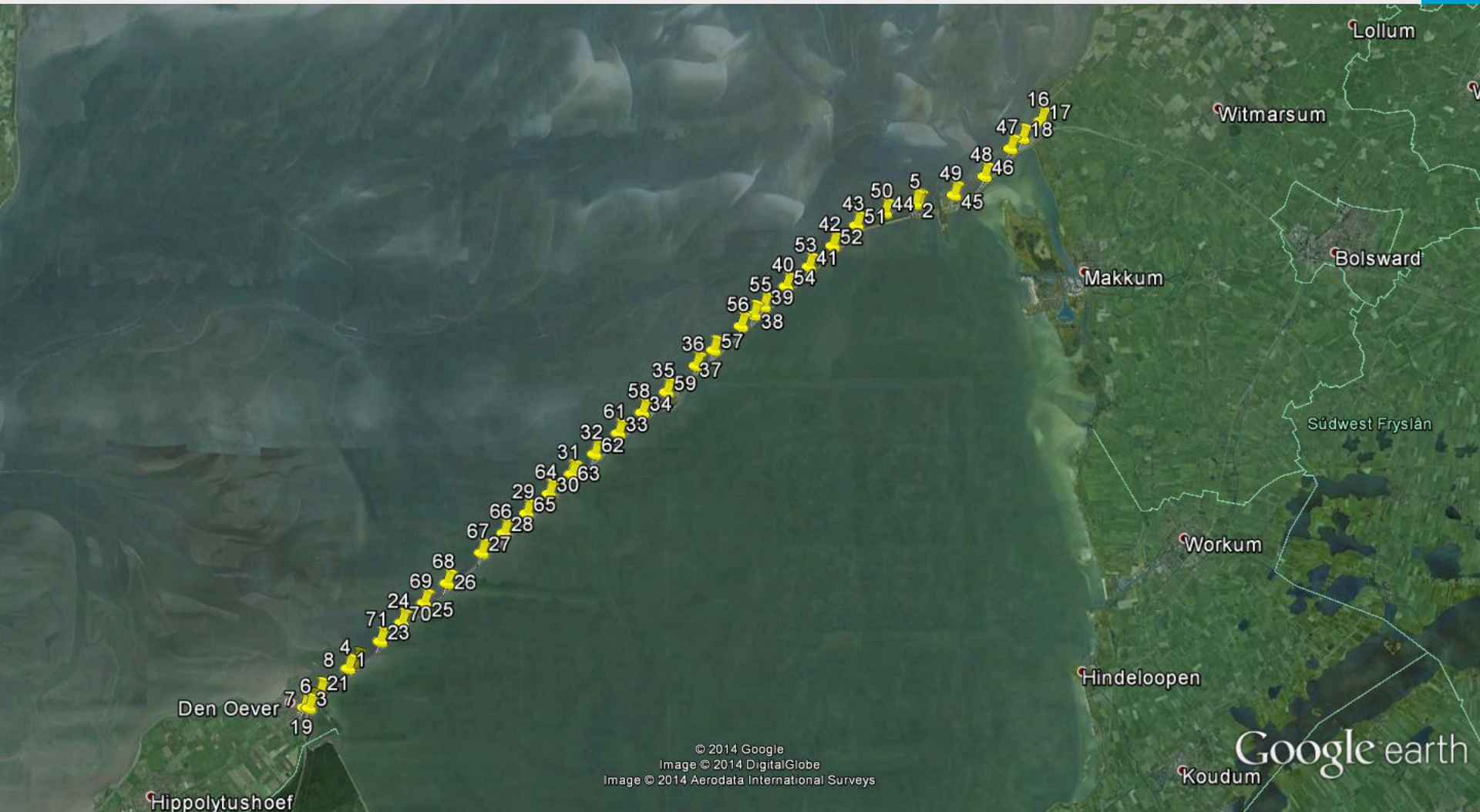
# Sensors in a road segment



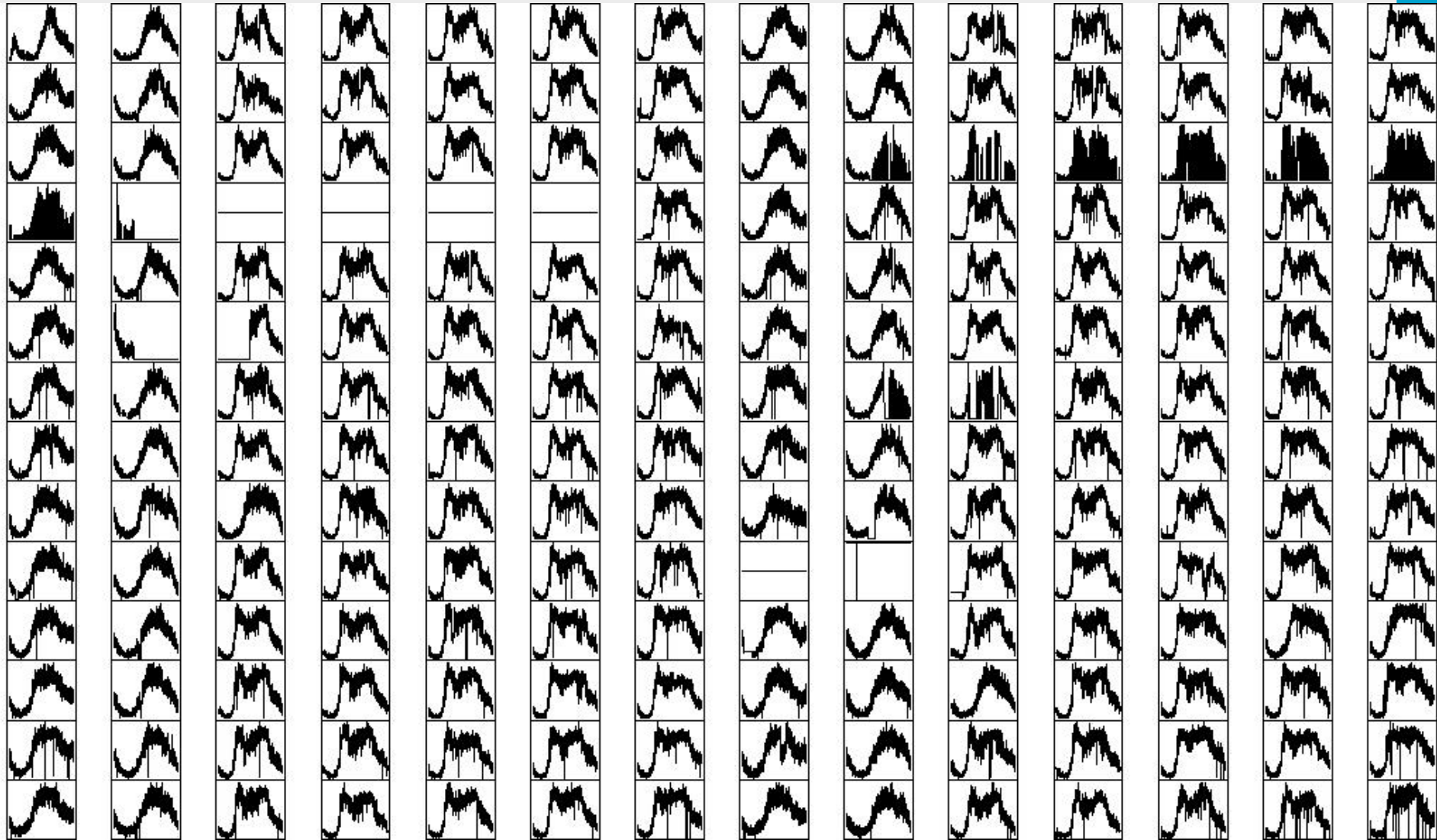
# A special dike



# Road sensors in the dike



# Minute data of one sensor for 196 days





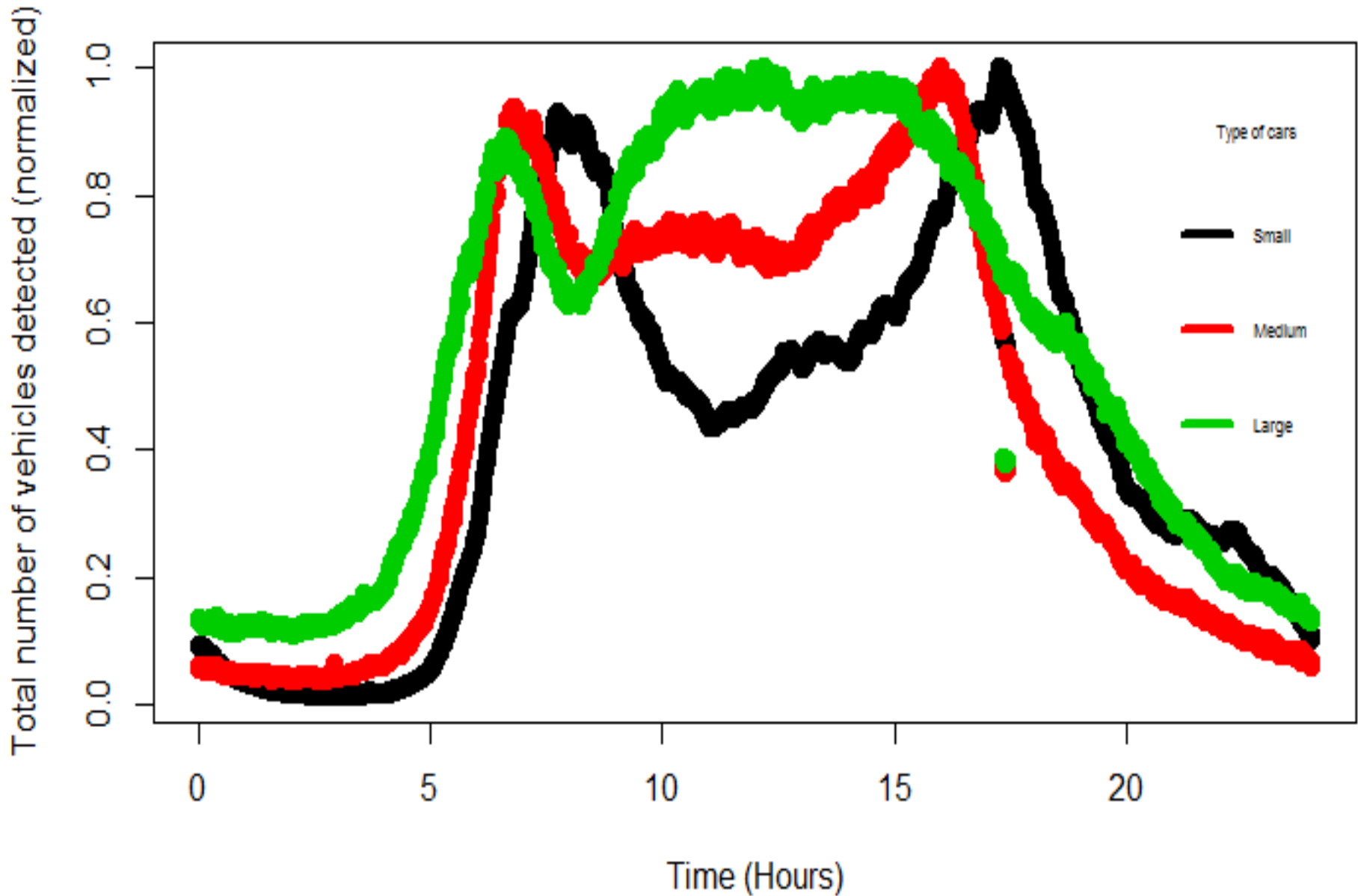
# Researching the data

Cross correlation between sensor pairs  
- Used to validate metadata

Trajectory speed vs. point speed  
- Average speed is 98 Km/h



# Small, medium-sized & large vehicles



# Issues and non-issues

## *Non-issues:*

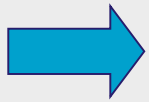
- Privacy
- Data acquisition

## *Issues:*

- Methodology
  - Selectivity
  - Quality
- Infrastructural needs
- Other issues
  - Skills needed



# Data options



## Historical database

- Request data via web interface
- Minute data for all highways
  - 48 variables, around 2.5 TB (Jan 2010-April 2014)
  - Data at a higher aggregation level is edited

## Data stream

- Every minute, all data for **all** active sensors
- Has to be continuously collected

# Process of road sensor based statistics

- Select sensors on Dutch highways
- Preprocessing
  - Remove non-informative variables
  - Remove bad records
  - Calculate number of vehicles (per minute)
  - Quality indicators for daily data per sensor
- Dimension reduction of daily data
  - Exclude bad sensors
  - Reduce dimensions on same road and region
  - Obtain number of vehicles for each road and region
- Calculate traffic index
  - Calculate indices per region

# Conclusions

- Invest in methodological research and play with the data to get a grip on quality
- Traffic loop data are an ideal Big Data source
- Develop good relations with the data provider

