



HAL
open science

Exploiting traffic data

Jan Ramon

► **To cite this version:**

| Jan Ramon. Exploiting traffic data. 2018. hal-01879941

HAL Id: hal-01879941

<https://inria.hal.science/hal-01879941>

Submitted on 24 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploiting traffic data

Jan Ramon

INRIA Meet up

Lille, 14/12/2017



Contents

- Introduction
- The ADEME MUST project
- Future perspectives
- Personal data protection

Lots of data

Car

- GPS localization
- Engine operation
- Driving behavior
- Car sensors (seat belts, windows, lights, ...)

Driver

- Demographics
- Social network
- Agenda

Environment

- Road network
- Meteo

ADEME -MUST objectives

1. What value can we get from (mainly) car data?
2. Can we address
 - Congestion
 - Pollution
 - Road security
 - ... ?

Xee

- Recruit drivers
- Collect data

INRIA

- Analyze data
- Interpret results

MUST project

Domain experts

- CEREMA
- i-Trans

Users

- Metropole Lille
- Industry

Progress

- Year 1: preparation
 - algorithm development
 - driver recruitment
- Data stream since October 2017
 - 118 drivers (growing to 500)
 - 116K road segments
 - Only considering Lille area

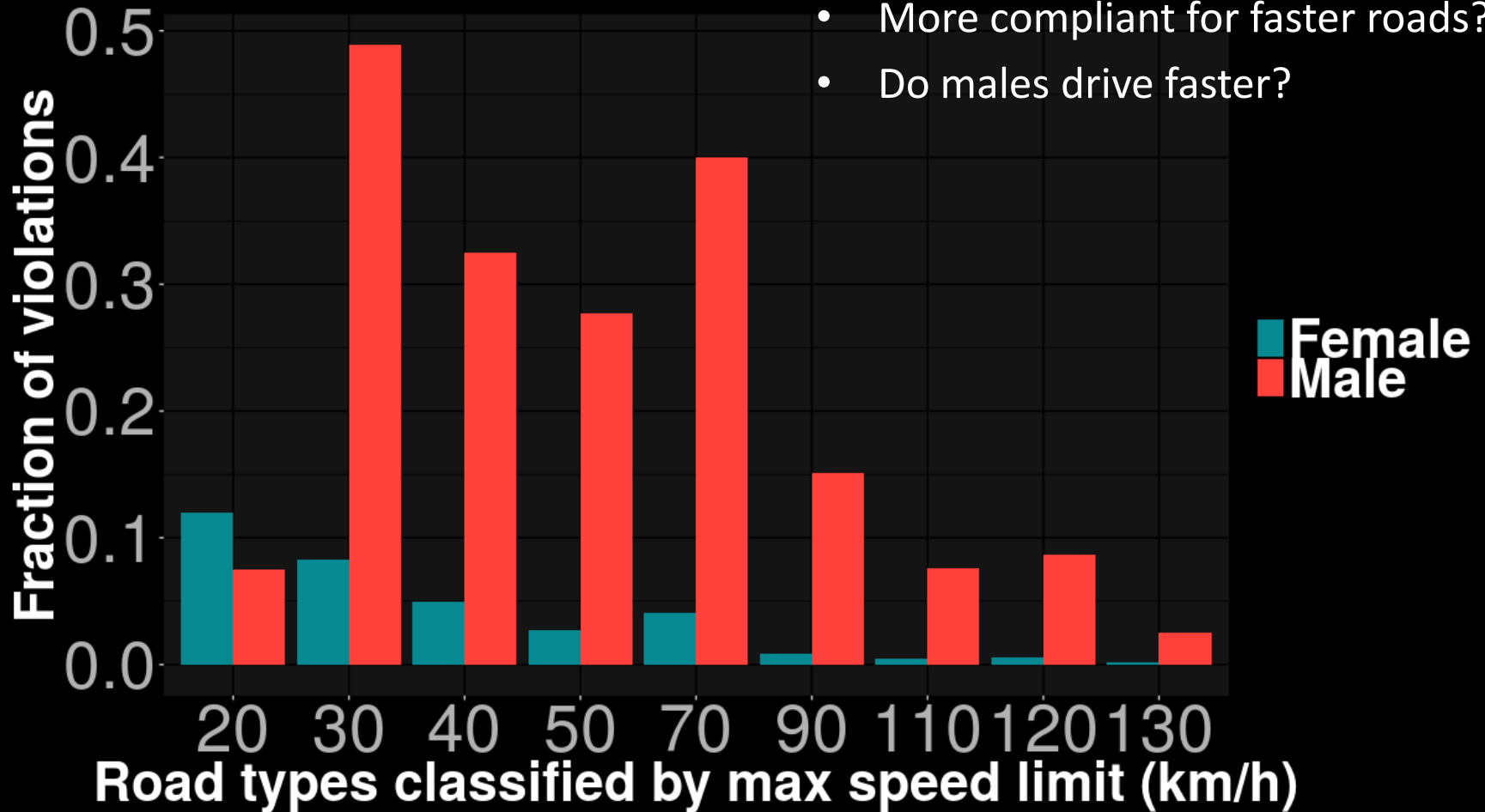


First analysis

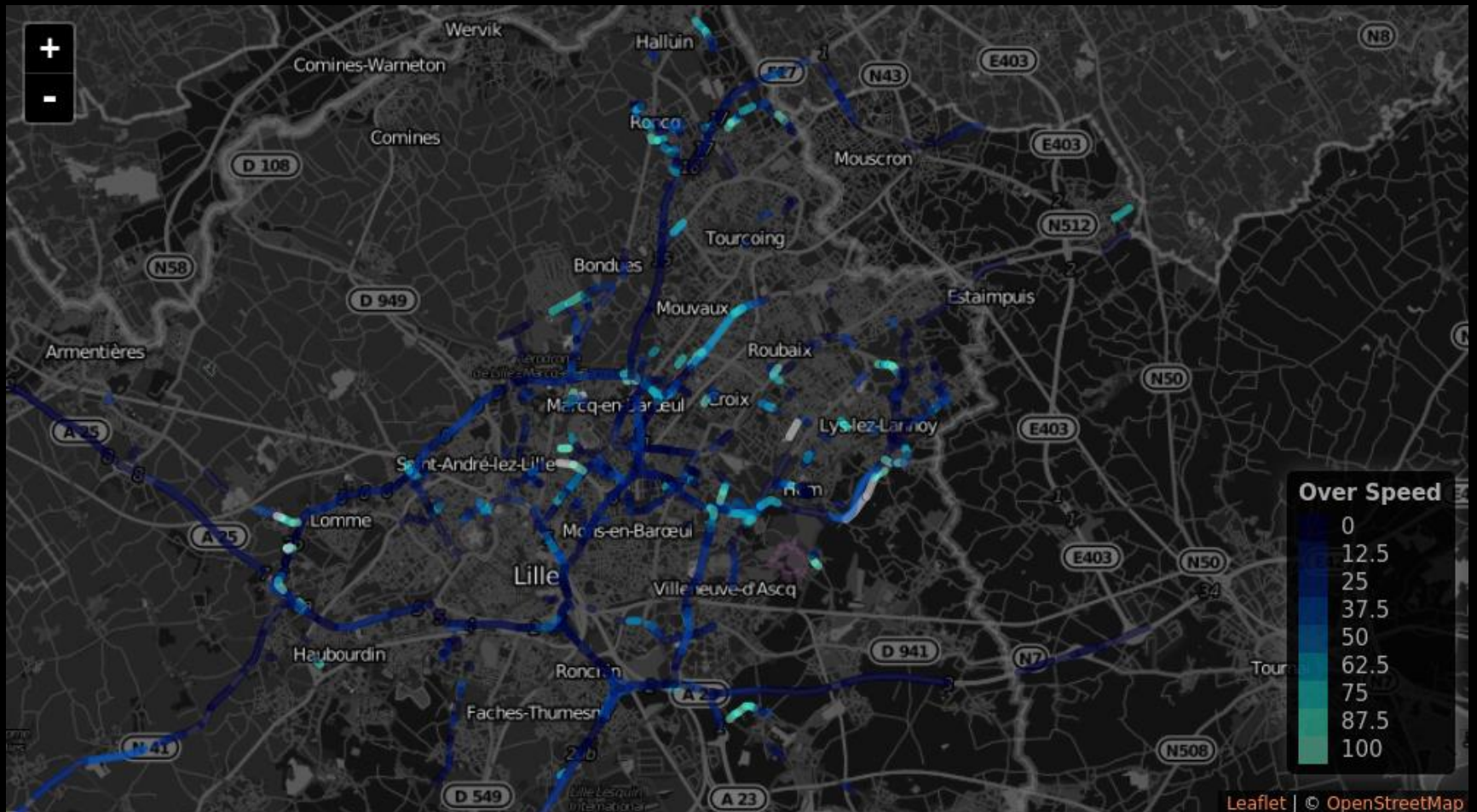
Age	Male	Female
18-25	8	1
25-35	32	4
35-45	26	3
45-55	7	0
55-65	6	2
65+	2	0
Unknown	27	0

Speeding

- Disclaimer: dataset noisy & small!
- More compliant for faster roads?
- Do males drive faster?



Uncleaned speeding on map



Current & next steps

- Driver behavior (acceleration, breaks, ...)
- Choice of trajectory
- Ride sharing
- Factors influencing congestions
- Fuel consumption (in practice)
- Parking (estimate duration, recommend tariffs, time to find free slot, ...)
- Effects of meteo


Future perspectives

- Data volume:
 - New cars will have more sensors (e.g. Camera)
 - Can generate 3Tb/day
- More cars will get “connected”
 - But they still may be far apart

Challenges

- Can't transfer data to a central place
 - 3Tb/day * 33M cars = 100Eb = 10^{20} b/day
- Personal and industrial data
 - GDPR : personal data should be protected
 - Even very innocent-looking data may leak information
 - Safety-critical data should be shared
 - Industries may not want to share some data

A decentralized approach

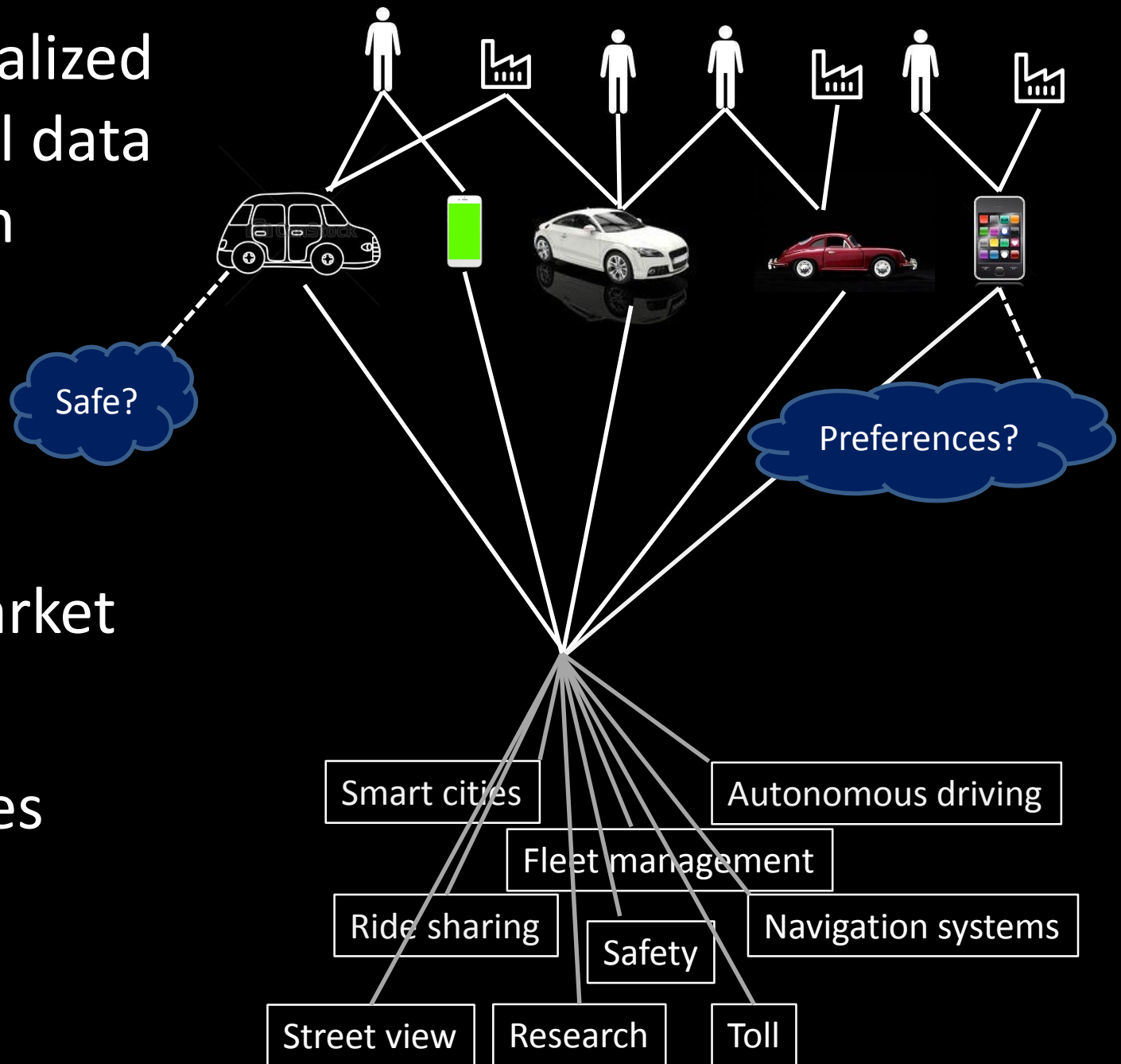
- Don't transfer data if not really needed
- Lower costs
 - Save on central storage
 - Save communication costs (or share only locally)
 - Exploit decentralized computational power
- More potential applications
 - Real-time
 - Perform tasks which are hard in a central HPC model (and even tasks where “move code to car” is not sufficient)
 - E.g. Street view updating
- Lower risks
 - More control on personal/industrial data use by citizen/OEM
 - Data not collected centrally can't leak/reveal information
- ERC  -PoC SOM project

- Decentralized personal data platform

- AI

- Data market

- Use cases



Conclusions

- ADEME-MUST: (wider) exploration of exploitation of mobility data
- Preparing for future challenges
- Lots of possible application, protection of data needed.

Questions?

