1. About movement data

2. A flashback to the past

3. Mobility data analytics pipeline

4. A real-world use case

Types of mobility data analytics

- Discovering groups and outliers
 - Background task: calculating similarity between (sub-)trajectories
- Detecting frequent routes (hot paths) and frequent locations (hot spots)
 - and their in-frequent counterparts
- Trajectory prediction tasks, etc.









Big Mobility Data Analytics Challenges

Volume **V**elocity



12K distinct ships/day, 200M AIS signals/month in EU waters



Historical & aggregated data, geographical & environmental data, contextual data, etc.

Veracity



Noisy and error-prone data due to receivers limited coverage, positioning devices switch-off

Image source: (Claramunt et al. 2017)

<u>V</u>ariety

Getting familiar with the data



Getting familiar with the data (cont.)

ADS-B and IFS-radar aviation datasets

- ADS-B: aircraft positions over Europe during 1 day; each aircraft reports, on average, every 28 sec.
- IFS-radar: aircraft positions over Spain during 1 week; each aircraft is captured, on average, every 5 sec.





GPS data pre-processing

- GPS recordings are distinct samples (p_i, t_i) of our movement
- Data pre-processing tasks:
 - Cleansing (noise removal, smoothing, map matching, etc.)
 - Transformation (trajectory segmentation, simplification, resampling, etc.)
 - Enrichment (semantic annotation, data fusion, etc.)



Data pre-processing (cont.)





Trajectory segmentation

 Goal: Segment sequences of points in homogeneous sub-sequences (= trajectories)



- Identification via stops or raw (spatial / temporal) gap
- Identification via prior knowledge (e.g. office hours, sleeping hours)
- Correlation-based identification (ideas from time-series segmentation)



Trajectory simplification & resampling

- The need for simplification: efficiency in storage, processing time, etc.
 - Goal: maintain the original 'signature' as much as possible by keeping a set of critical points only
 - Offline (Meratnia & de By, 2004) vs. online methods (Patroumpas et al. 2015; 2017)
- The need for resampling (Georgiou, 2017): some analytic methods assume fixed and aligned steps in location timing, unrealistic in GPS world
 - Issues: dense vs. sparse sampling? Offline vs. online process?

