

## Stand May 2, 2018:

### Schlechte Daten

- keine ID's: kann keine zwei Movements von einer Person aggregieren

⇒ Einteilung in Gruppen nicht wirklich Personen zuzuordnen

- eine Person ist morgens in anderem Cluster wie abends
  
- Möglich wäre es Movements basierend auf: Alter, Geschlecht, Strata ⇒ ID?
- Unterscheidung von zwei zeitgleichen Movements, welche demnach nicht der selben Person gehören können
- Problem: false Positives: wenn zwei unterschiedliche Personen als eine erkannt werden
- Mögliche Lösung des Problems:
  - (Möglicherweise als Datenbank-JOIN)
  - Als Join in Rapidminer:  $Alter + 126 \cdot Geschlecht \cdot Strata$
  - (als Erweiterung im Python Script)
- ID als Penalty für "schlechtes" Clustering

### Deliverables

- ein Plotter für alle möglichen Variable-Kombinationen und auch so etwas wie: DIST/TIME
- Python Datensatz Verringerung