Utrecht University

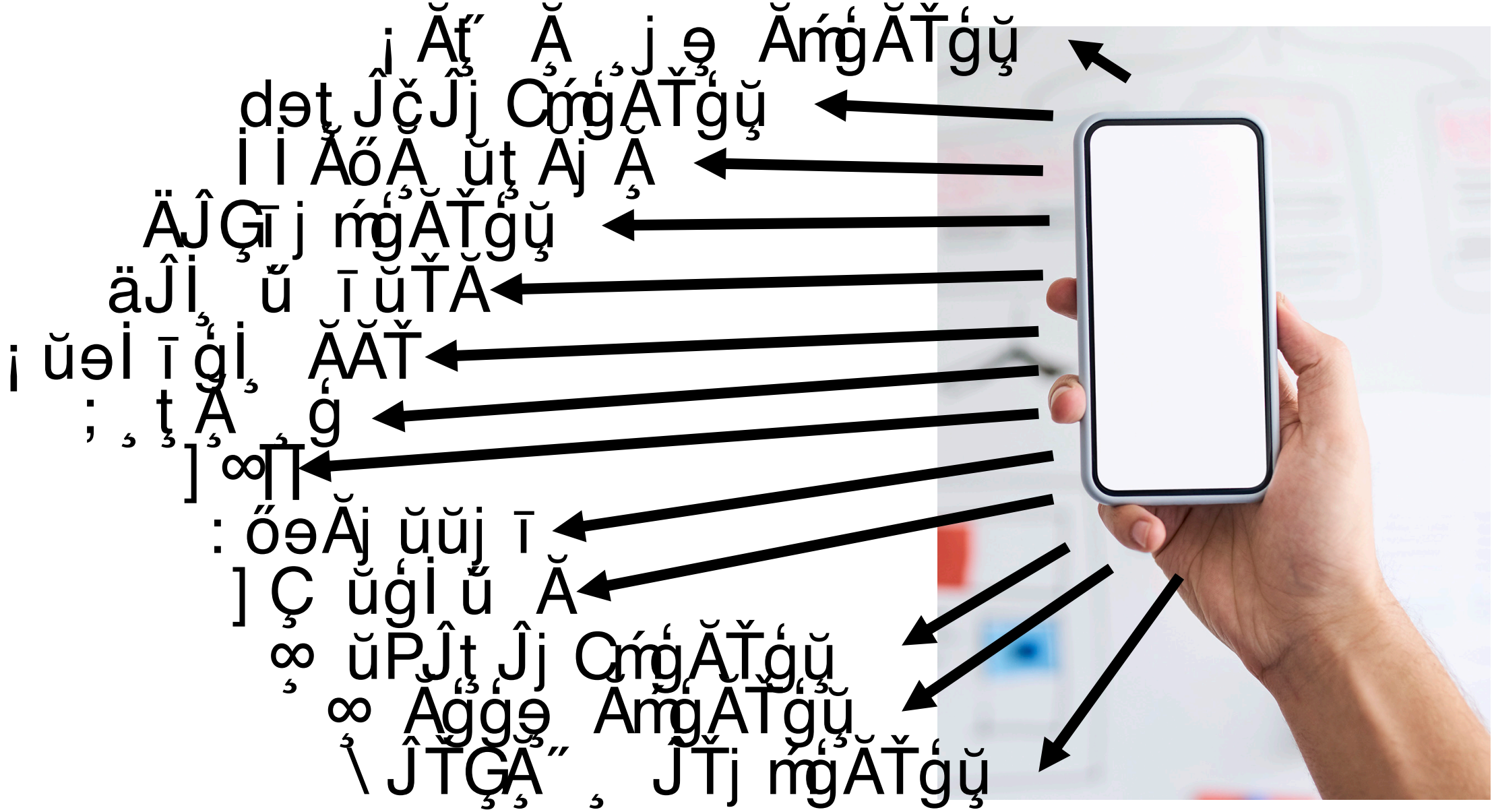
Centraal Bureau voor de Statistiek

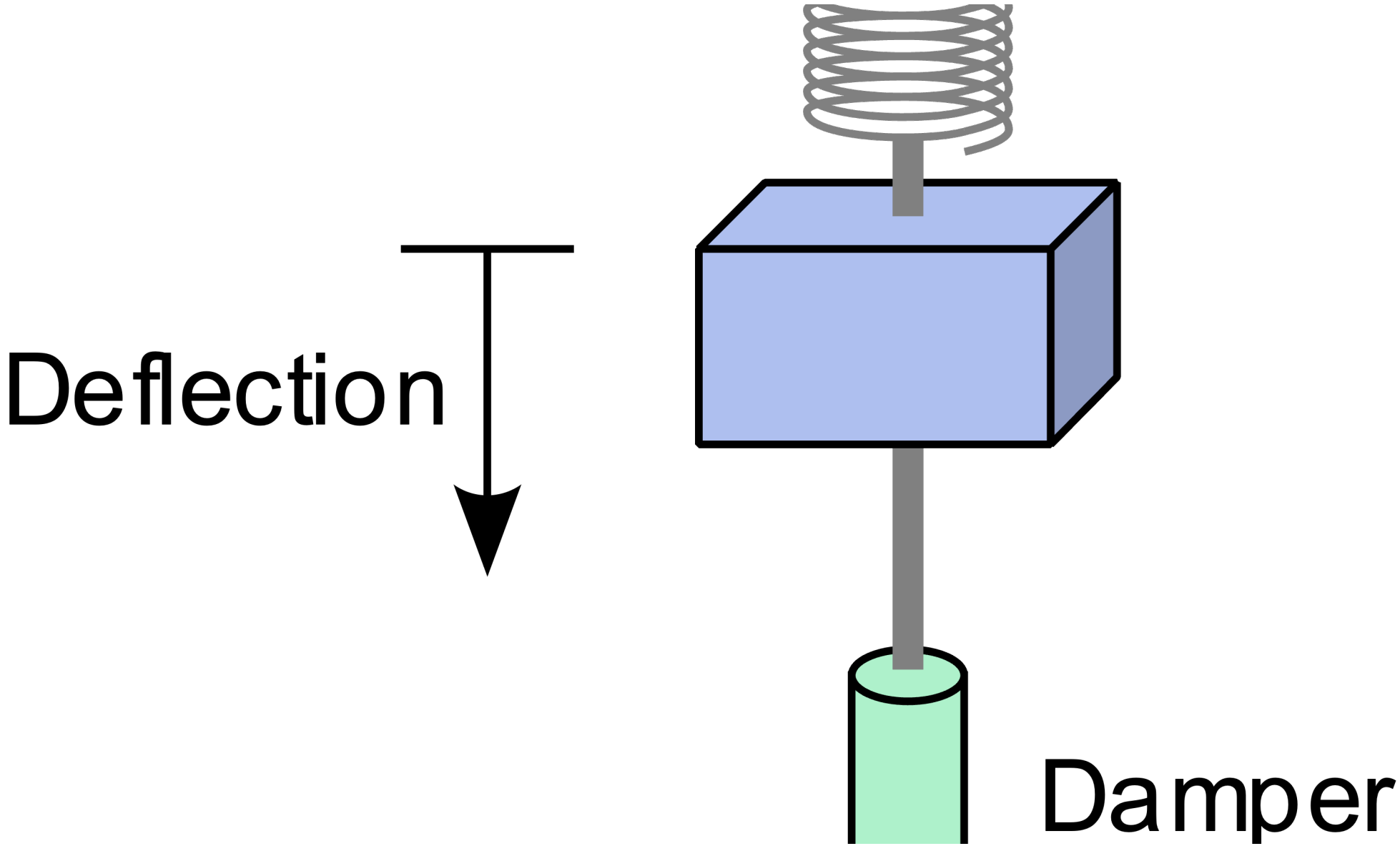
Smart surveys. Integrating sensors into surveys during data collection

Danielle McCool & Peter Lugtig

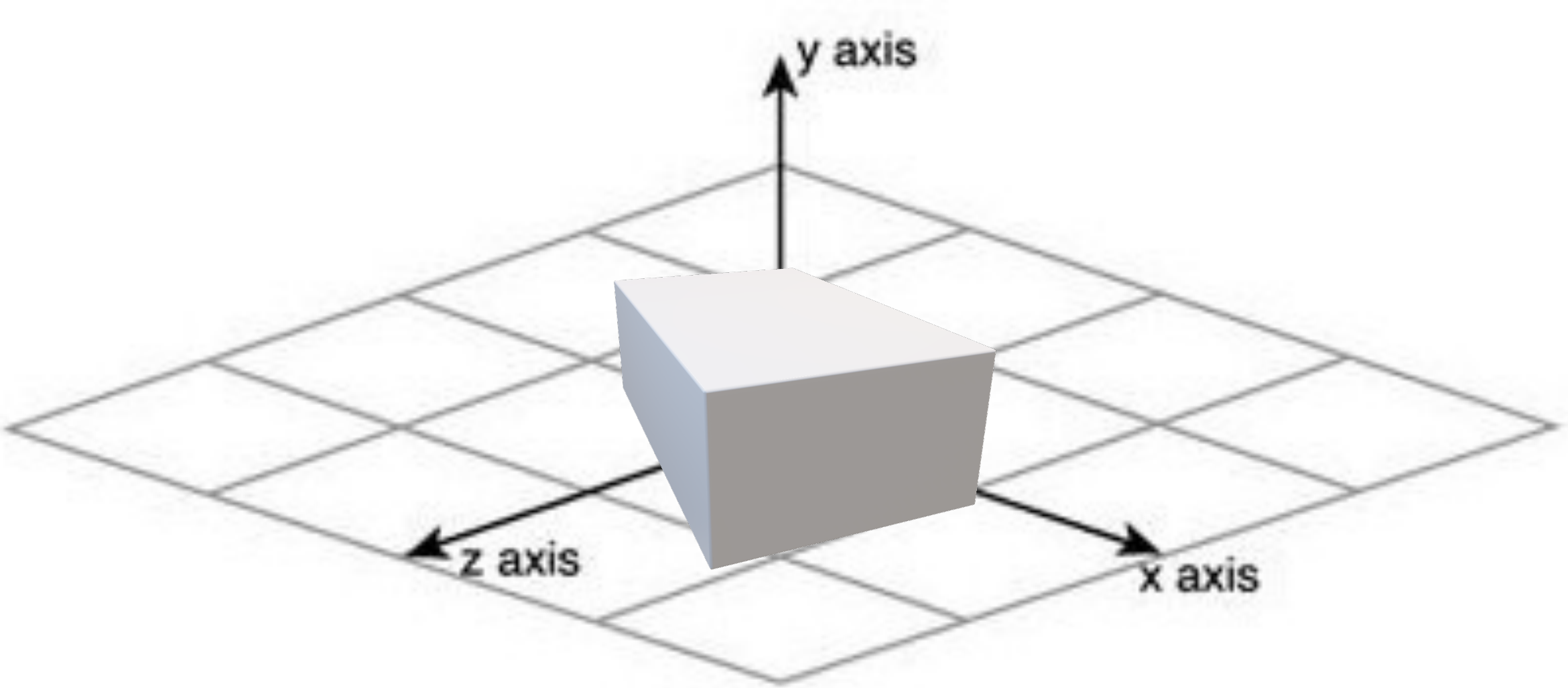
SIG Sensors

The promise: better data

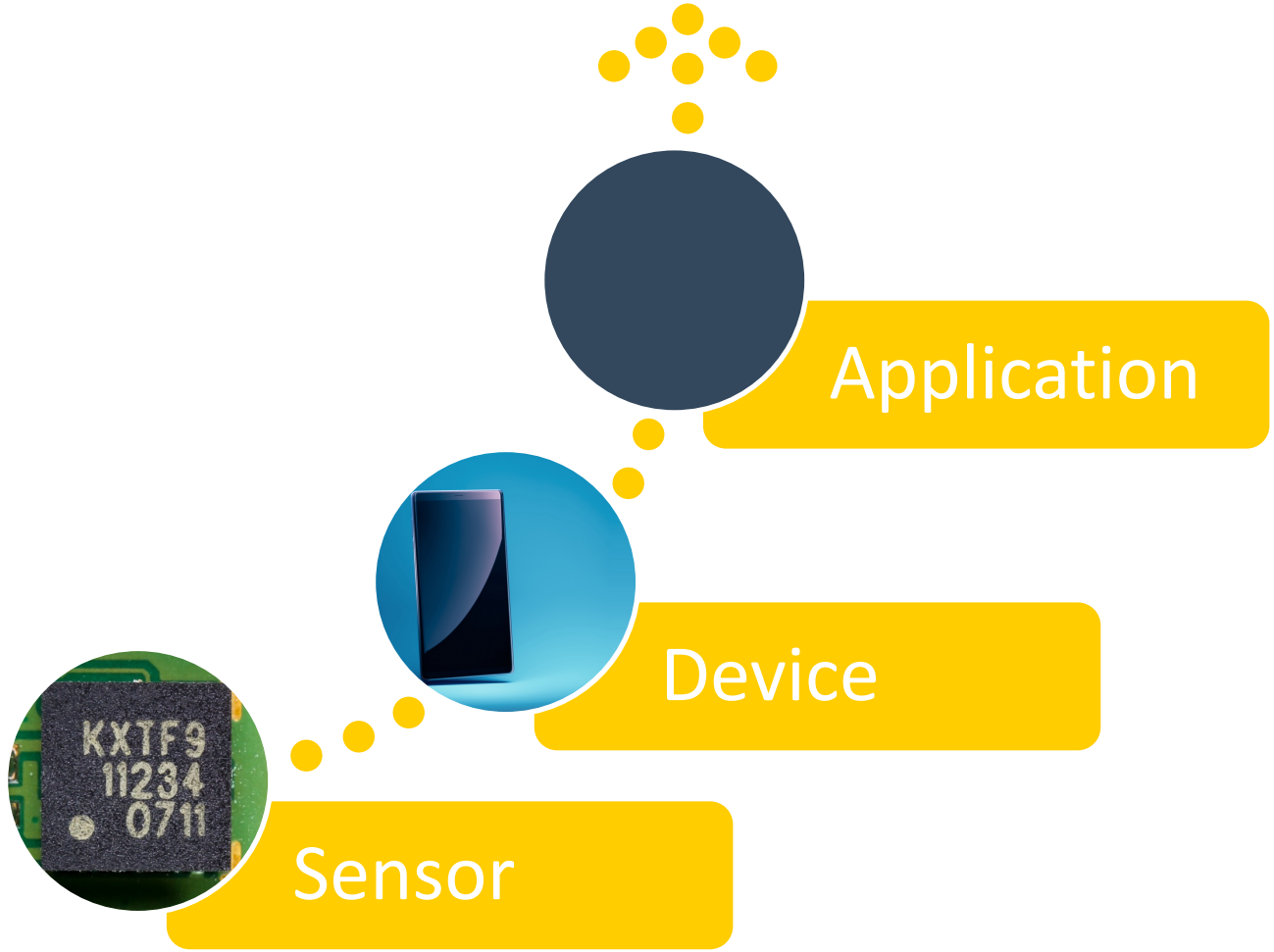


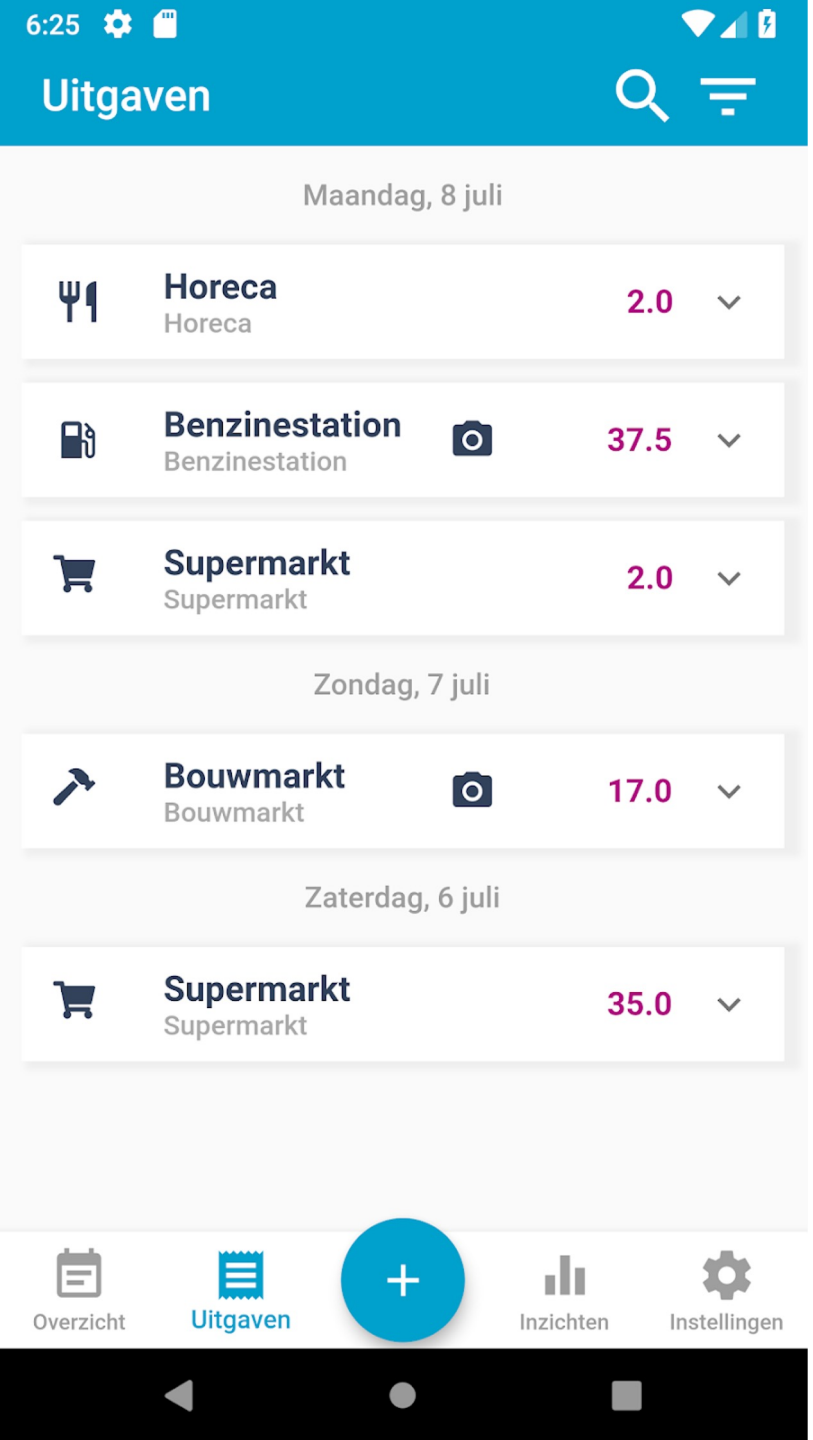


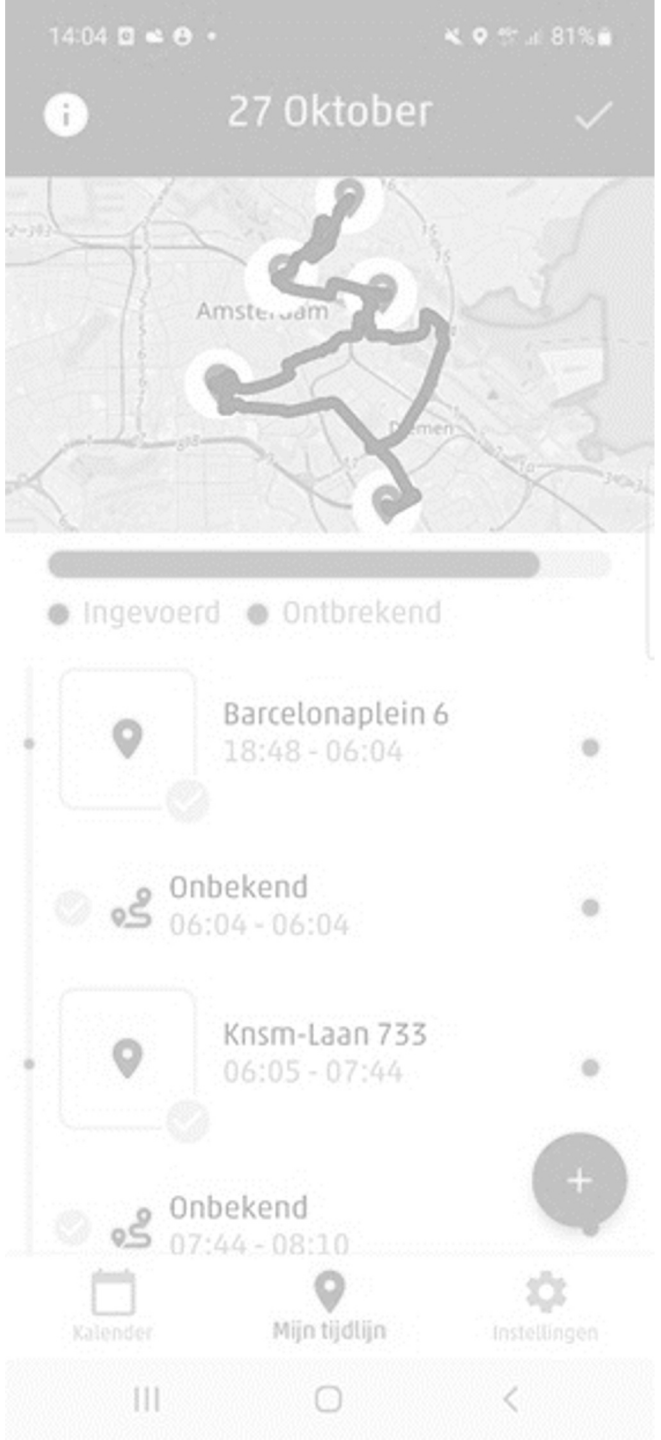
Traced by User:Stannered, Public domain, via Wikimedia Commons



Sensors aren’t (generally) designed to answer our research questions

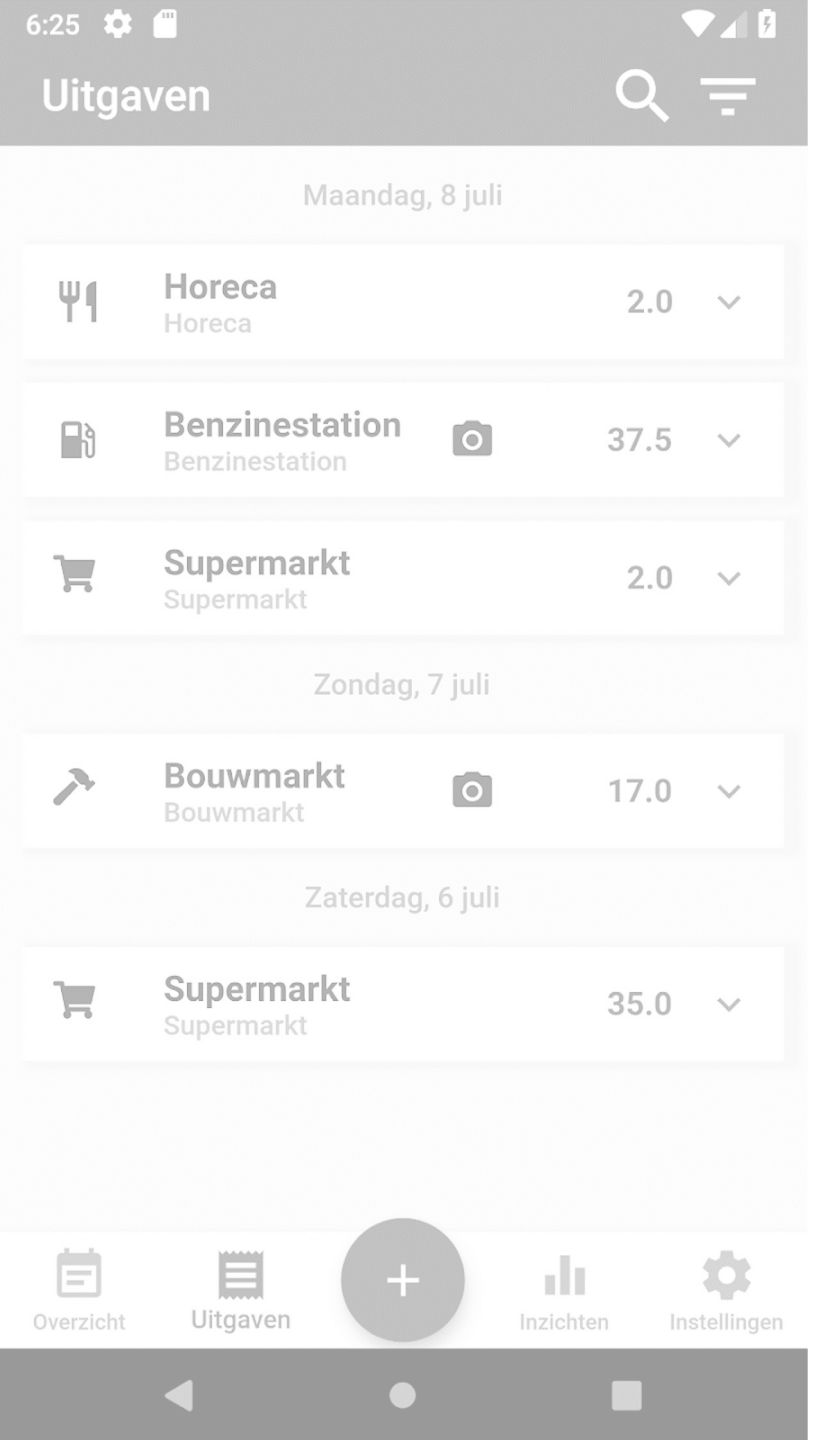


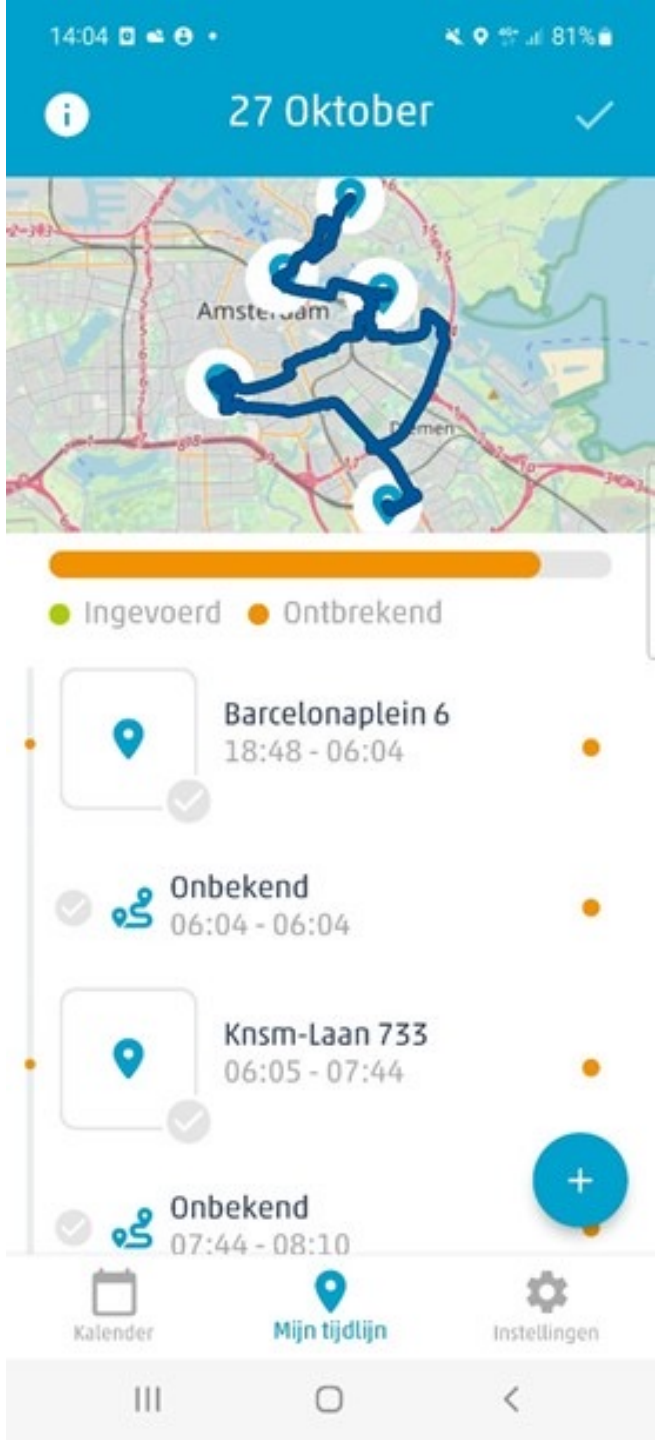




Current CBS app studies

- The app: Household Budget Survey (HBS)
- The sensor: Camera
- The goal: Reduce respondent burden in inputting purchases by scanning receipts





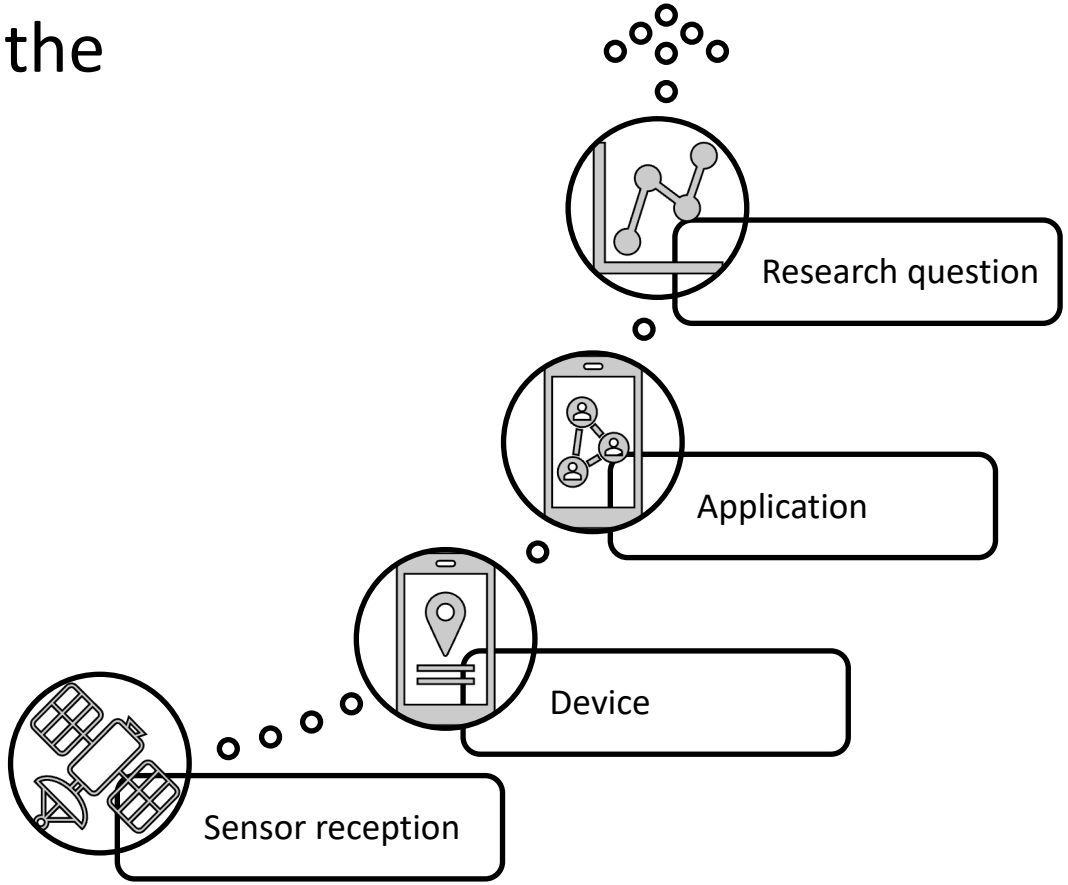
Current CBS app studies

- The app: Travel survey (AVA)
- The sensor: GPS/GNSS receiver
- The goal: Improve mobility reporting by filling in trip data

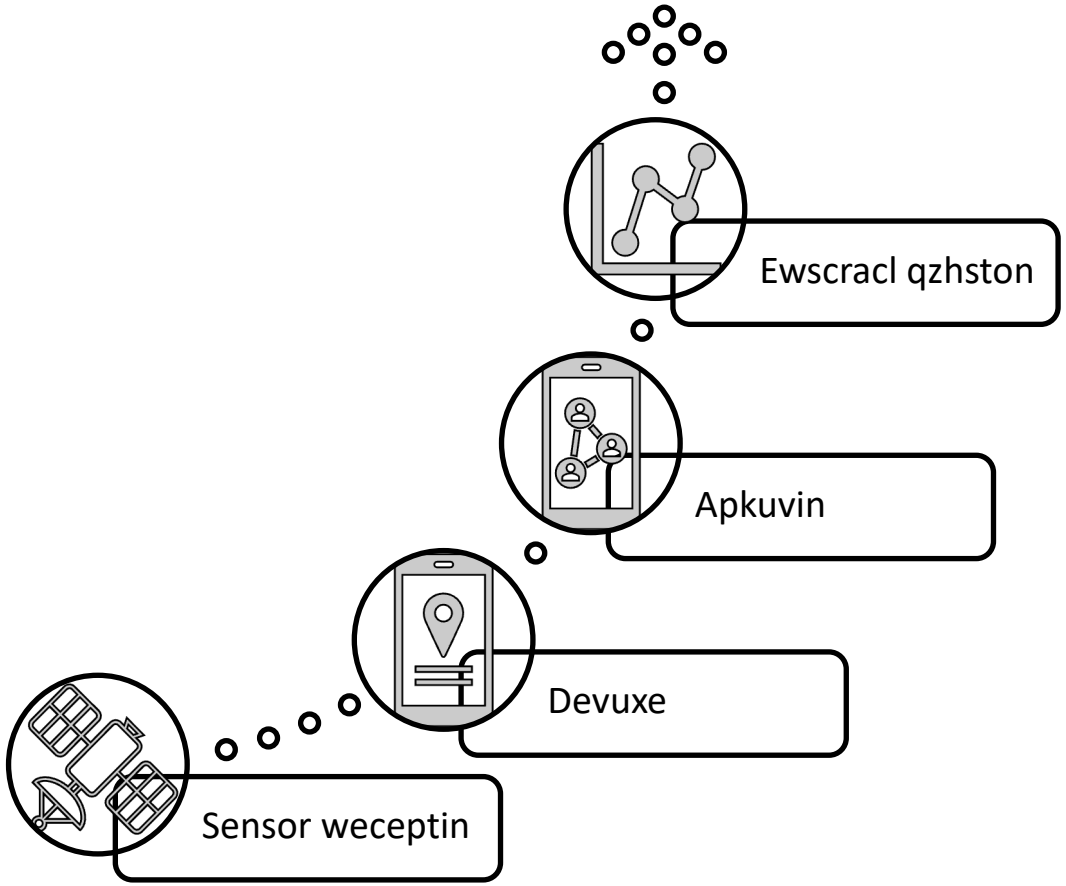
The promise: better data*

The promise: better, faster, easier data*

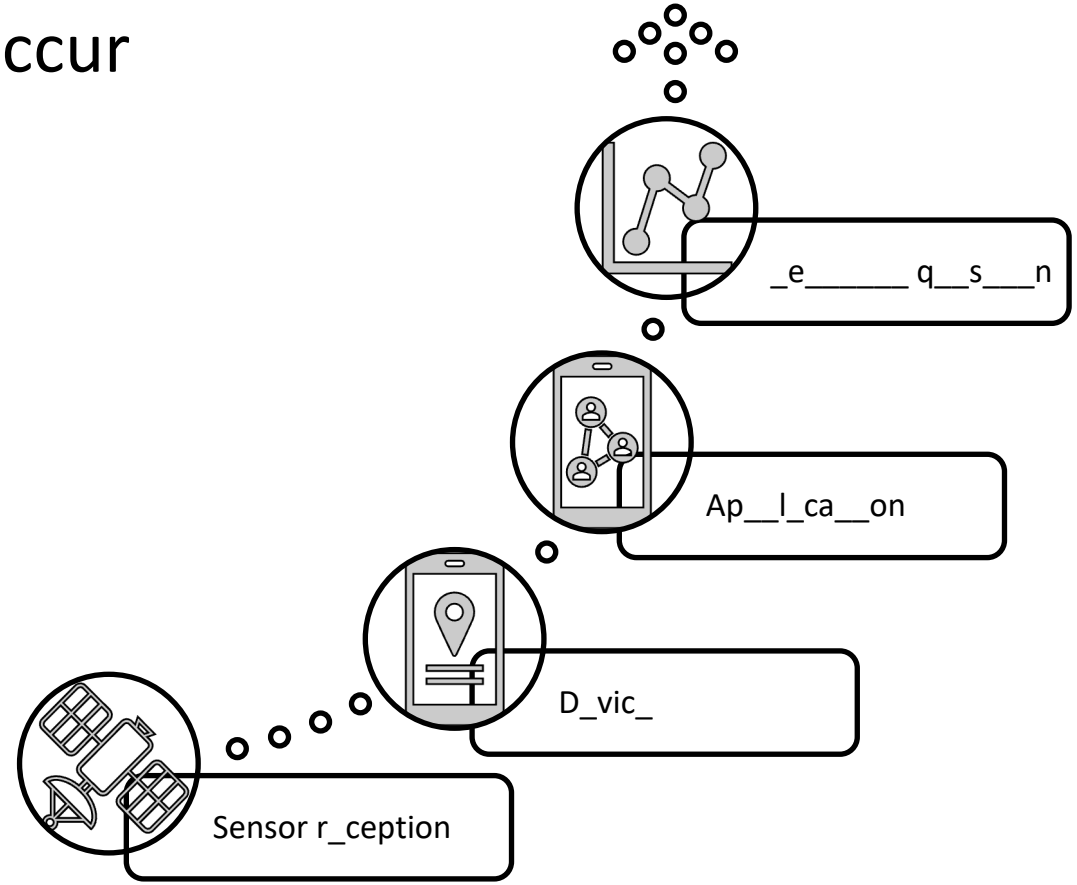
Example: Layers in the Travel App

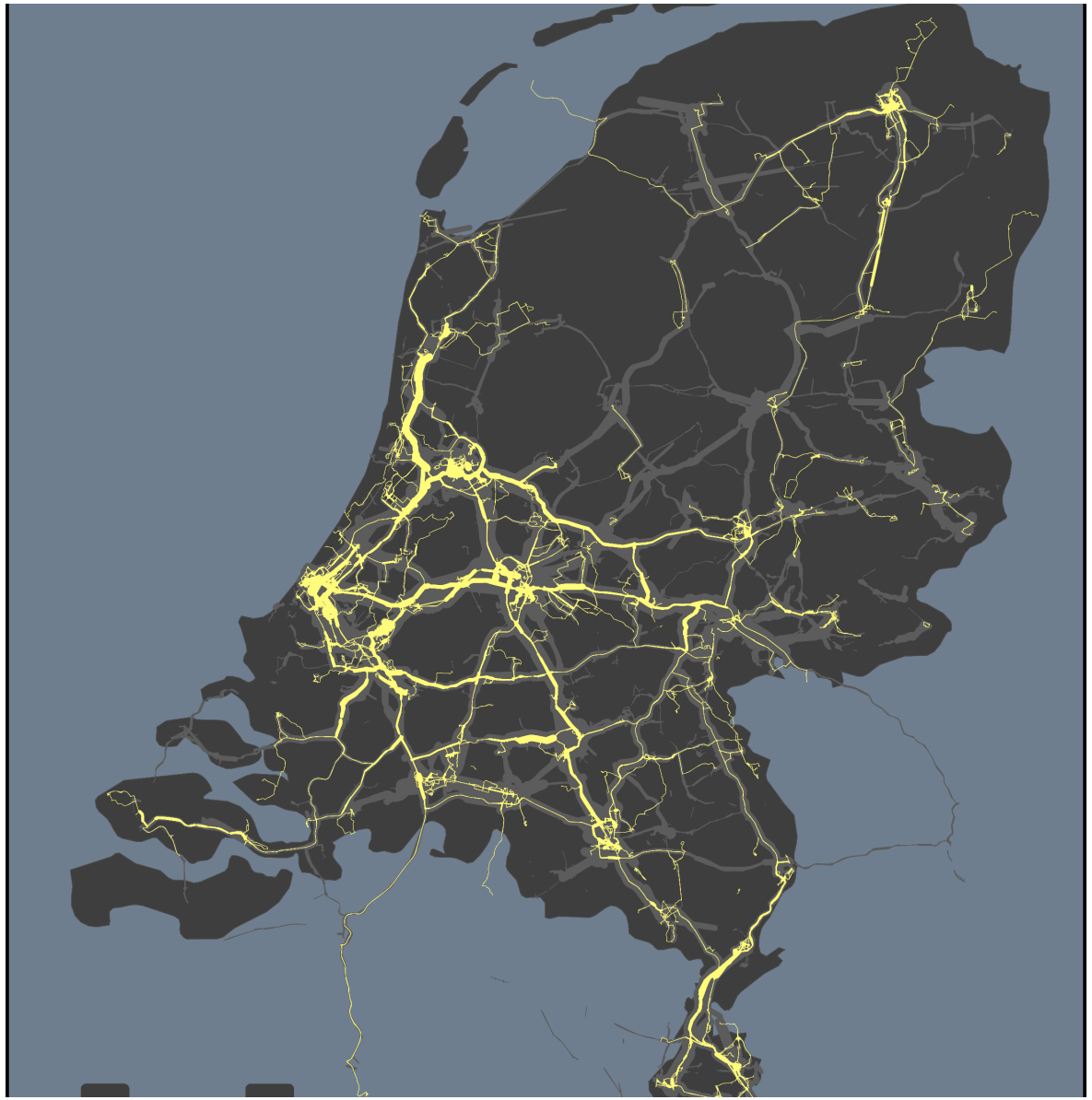


Wrong data can be reported at a layer



Missing data can occur at several layers

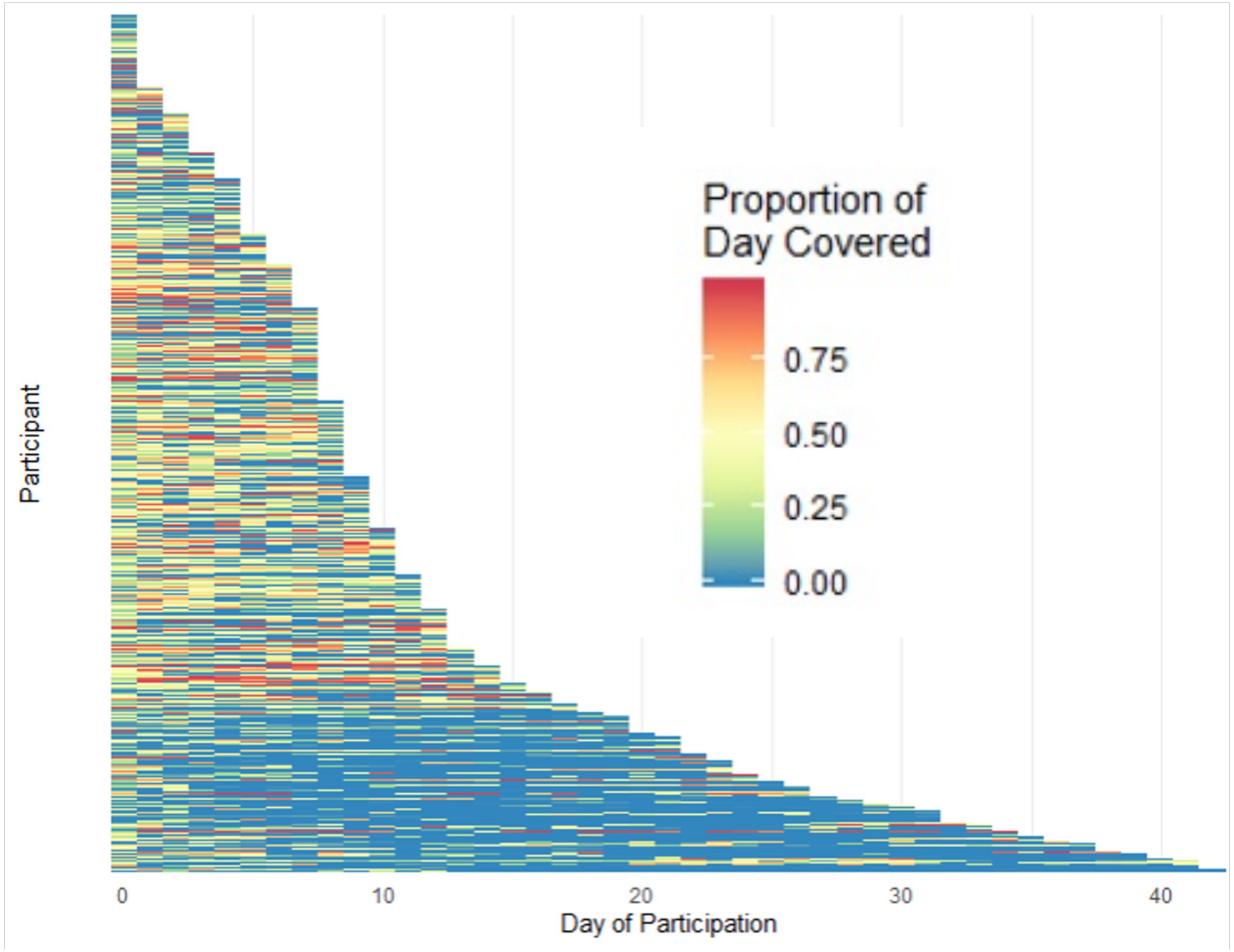




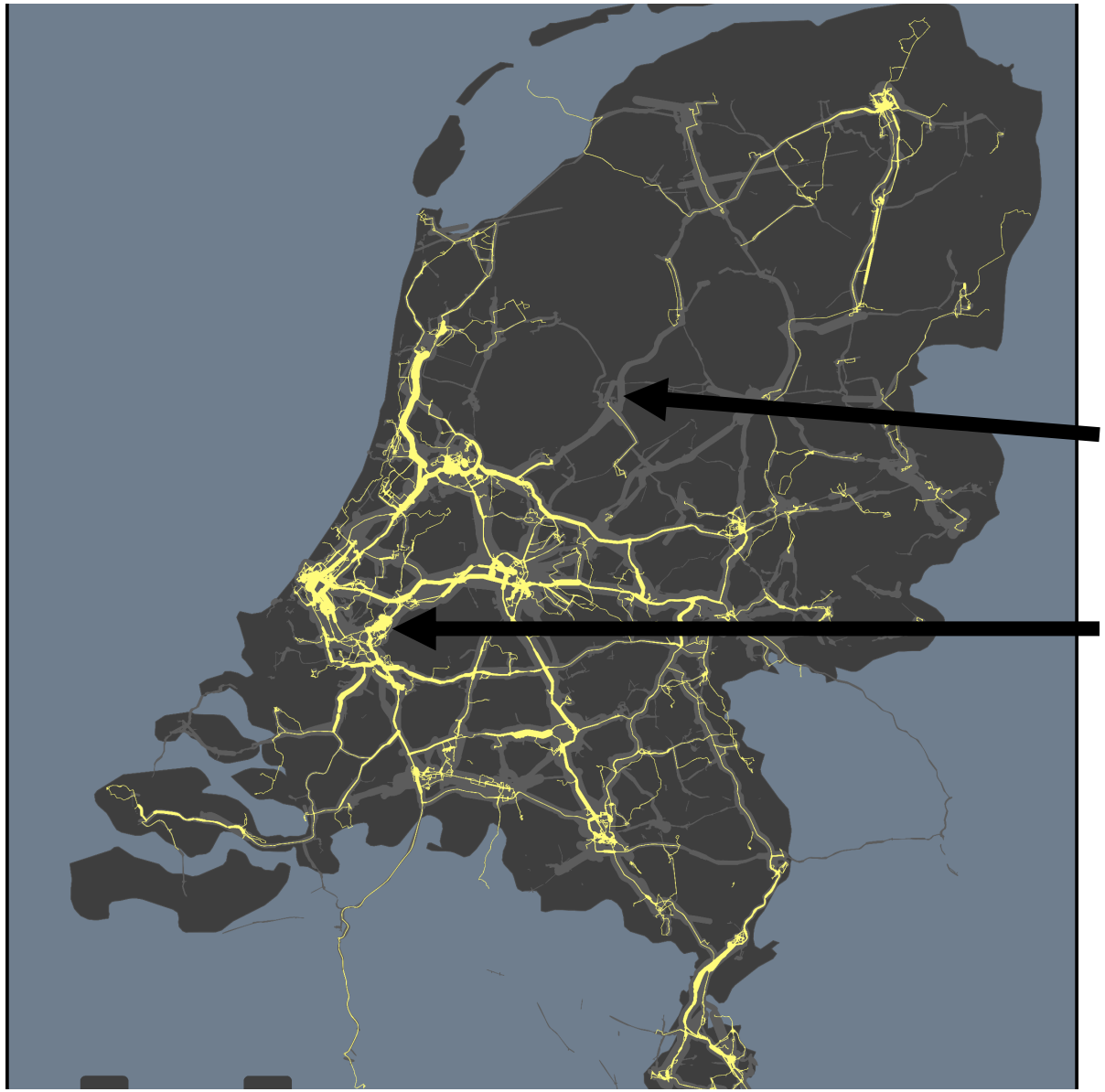
Lots of data

In total, over 2000 cumulative days of locations

Lots of holes



Average length of day recorded: ~ 10 hours

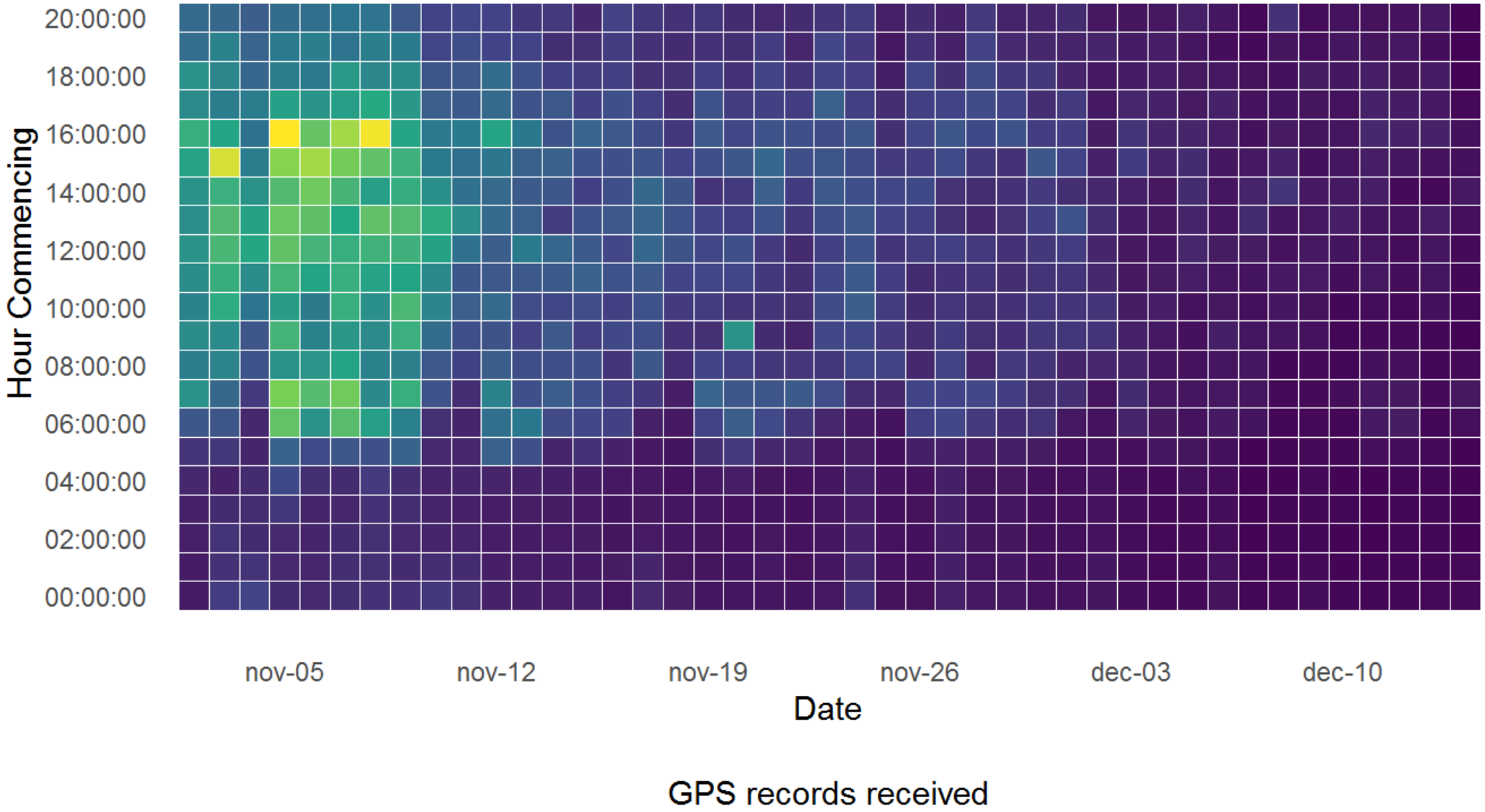


Lots of data

In total, over 2000 cumulative days of locations

Light grey: all routes traveled in the data

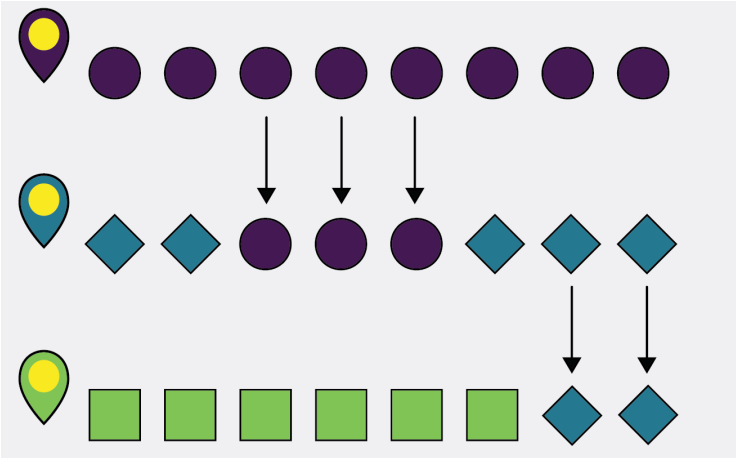
Yellow: routes traveled where the data are complete enough to use


Hour Commencing
20:00:00
18:00:00
16:00:00
14:00:00
12:00:00
10:00:00
08:00:00
06:00:00
04:00:00
02:00:00
00:00:00
nov-05
nov-12
nov-19
nov-26
dec-03
dec-10
Date
GPS records received

Imputation concept

Similar travel patterns can provide a basis for imputing long gaps in travel behavior.

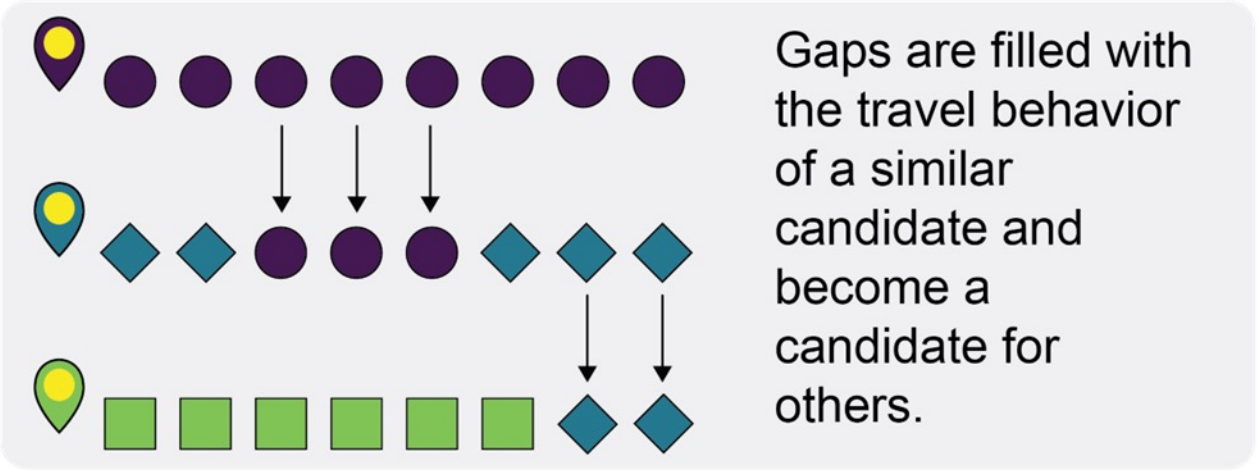
The imputation procedure

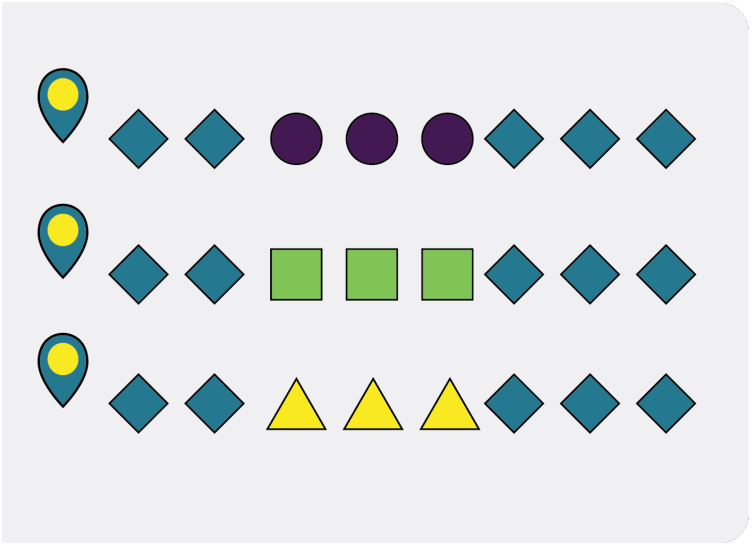


Gaps are filled with the travel behavior of a similar candidate.

These go on to become a candidate for others.

The imputation procedure





Using multiple candidates for filling each gap enables measurement of uncertainty and offers more robust estimates.

Imputation consequences (simulation)

Method comparison across all cases
Imputing travel distance

	Abs Bias	Med Bias
LI	5.9 Km	−0.3 Km
MI	1.9 Km	5.7 Km
TWI	1.1 Km	2.0 Km
DTWBI	1.8 Km	0.0 Km
DTWBMI-HI	0.7 Km	0.0 Km
DTWBMI-LO	0.6 Km	0.0 Km

Compared to linear interpolation (LI), and mean imputation (MI), the methods that imputed travel behavior using this method were less biased

At each layer another algorithm

Takeaways

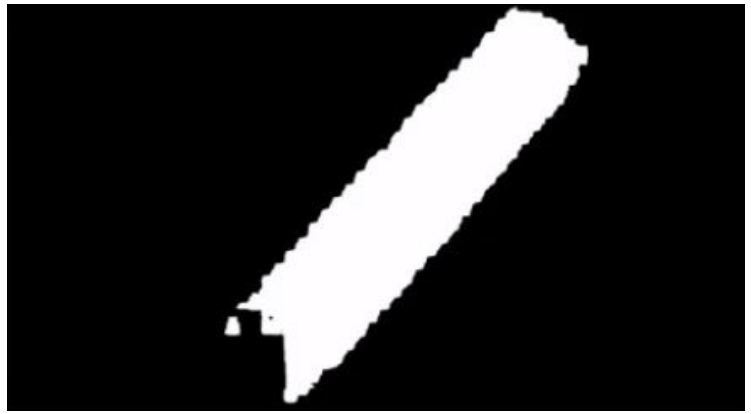
- Sensors are exciting
- How we get from the physical measurements to answerable research questions is important
- Decisions made at each layer have downstream impacts
- The data we capture *can* be better

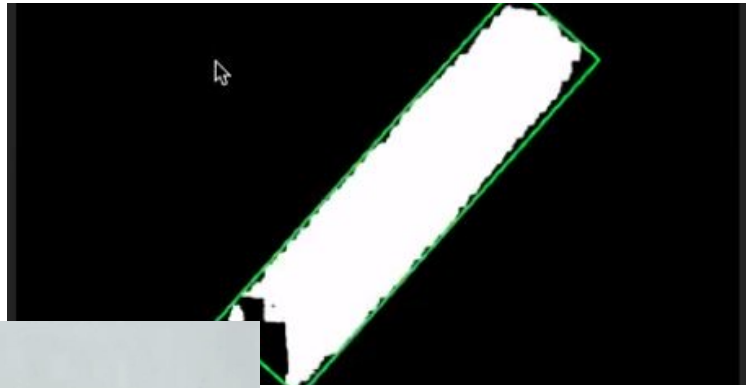
Utrecht University

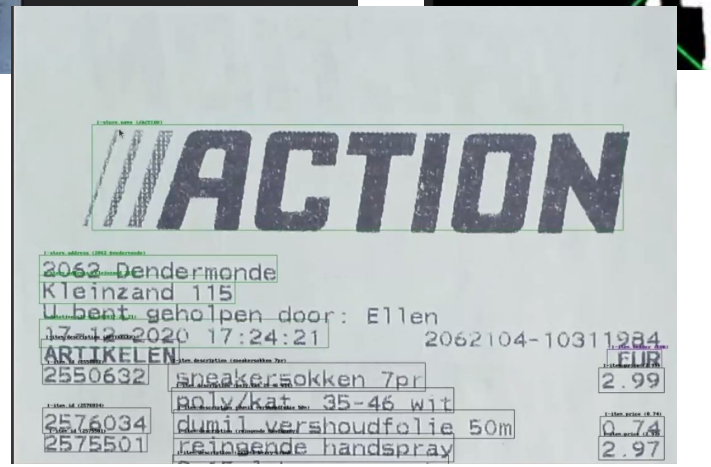
Sharing science,
shaping tomorrow

Apps Detail

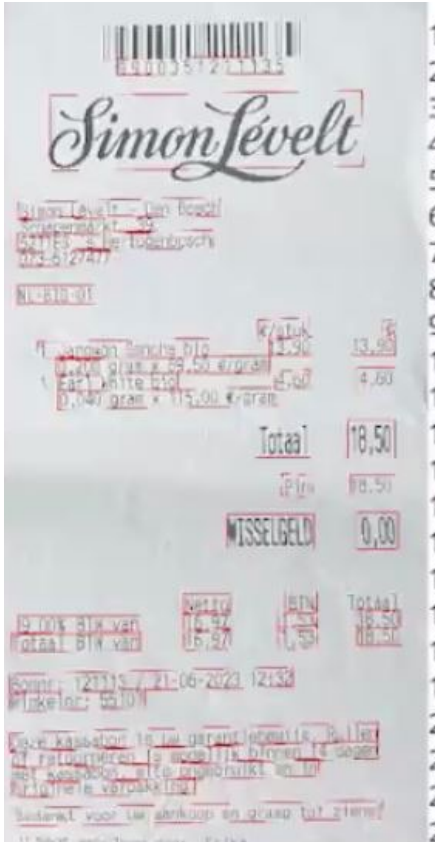
Budget Study Layers





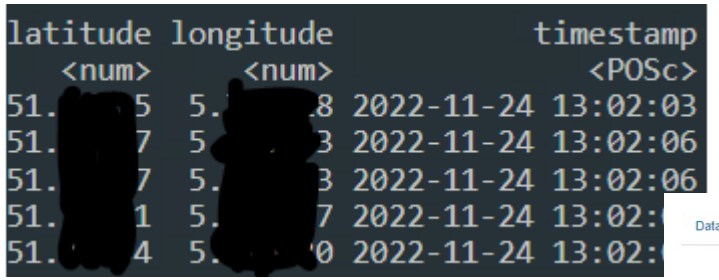


1. Get photo from device
2. Receipt segmentation
3. Bounding the rectangle
4. Cropping and rotating
5. OCR
6. ...?
7. Answers

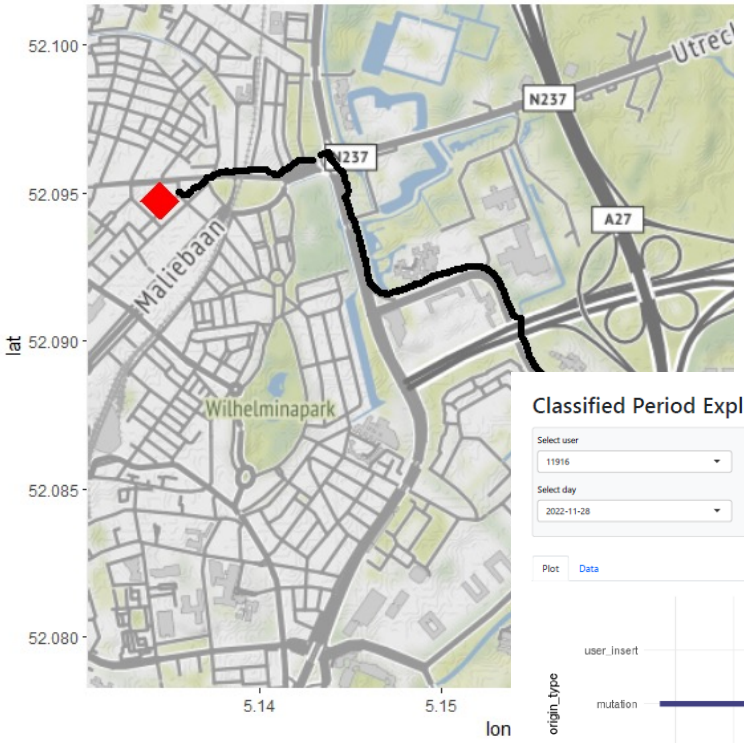


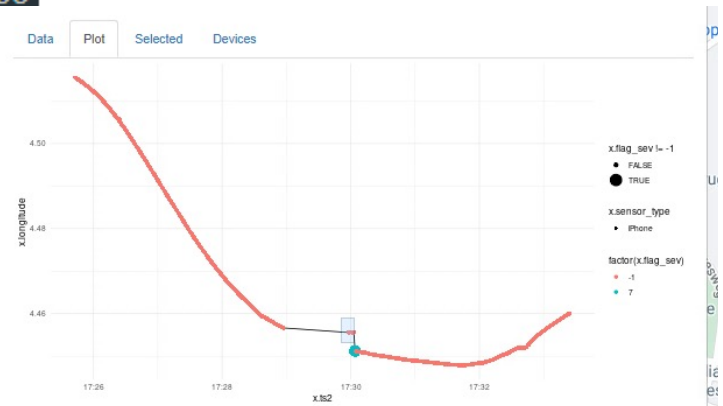
1: "9900351211135" 0.993
2: "Simonfevelt" 0.982
3: "Simon Levelt - Den Bosch" 0.958
4: "Schapenmarkt 39" 1.000
5: "5211ES s Hertogenbosch" 0.976
6: "073-6127477" 0.946
7: "NL-BI0-01" 0.953
8: "/stuk" 0.990
9: "" 0.999
10: "13,90" 0.955
11: "13,90" 0.998
12: "1 Jangwon Sencha bio" 0.992
13: "0,200 gram x 69,50 /gram" 0.985
14: "4,60" 0.991
15: "4,60" 0.997
16: "Earl White bio" 0.969
17: "0,040 gram x 115,00 /gram" 0.957
18: "Totaal" 0.994
19: "18,50" 0.999
20: "Pin" 0.999
21: "18,50" 0.999
22: "WISSELGELD" 0.893
23: "0,00" 0.972
24: "BTW" 0.985
25: "Totaal" 0.999
26: "Netto" 0.999
27: "1,53" 0.901
28: "18,50" 0.972
29: "16,97" 0.996
30: "9,00% BTW van" 0.992
31: "18,50" 0.943
32: "1,53" 0.998
33: "16,97" 0.997
34: "Totaal BTW van" 0.968
35: "Bonnr: 121113/ 21-06-2023 12:32" 0.977
36: "Winkelnr: 55107" 0.994
37: "Deze kassabon is uw garantiebewijs. Ruilen" 0.994
38: "of retourneren is mogelijk binnen 14 dagen" 0.988
39: "met kassabon, mits ongebruikt en in" 0.994
40: "originele verpakking." 0.999
41: "Bedankt voor uw aankoop en graag tot ziens!" 0.987

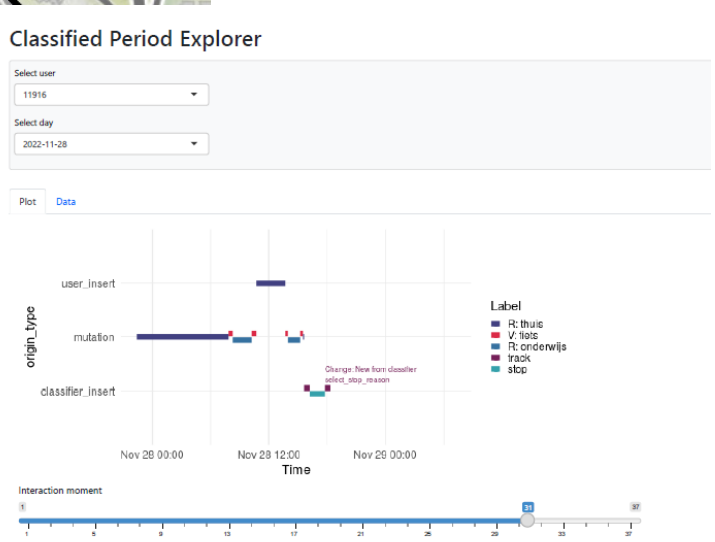
Mobility Study Layers



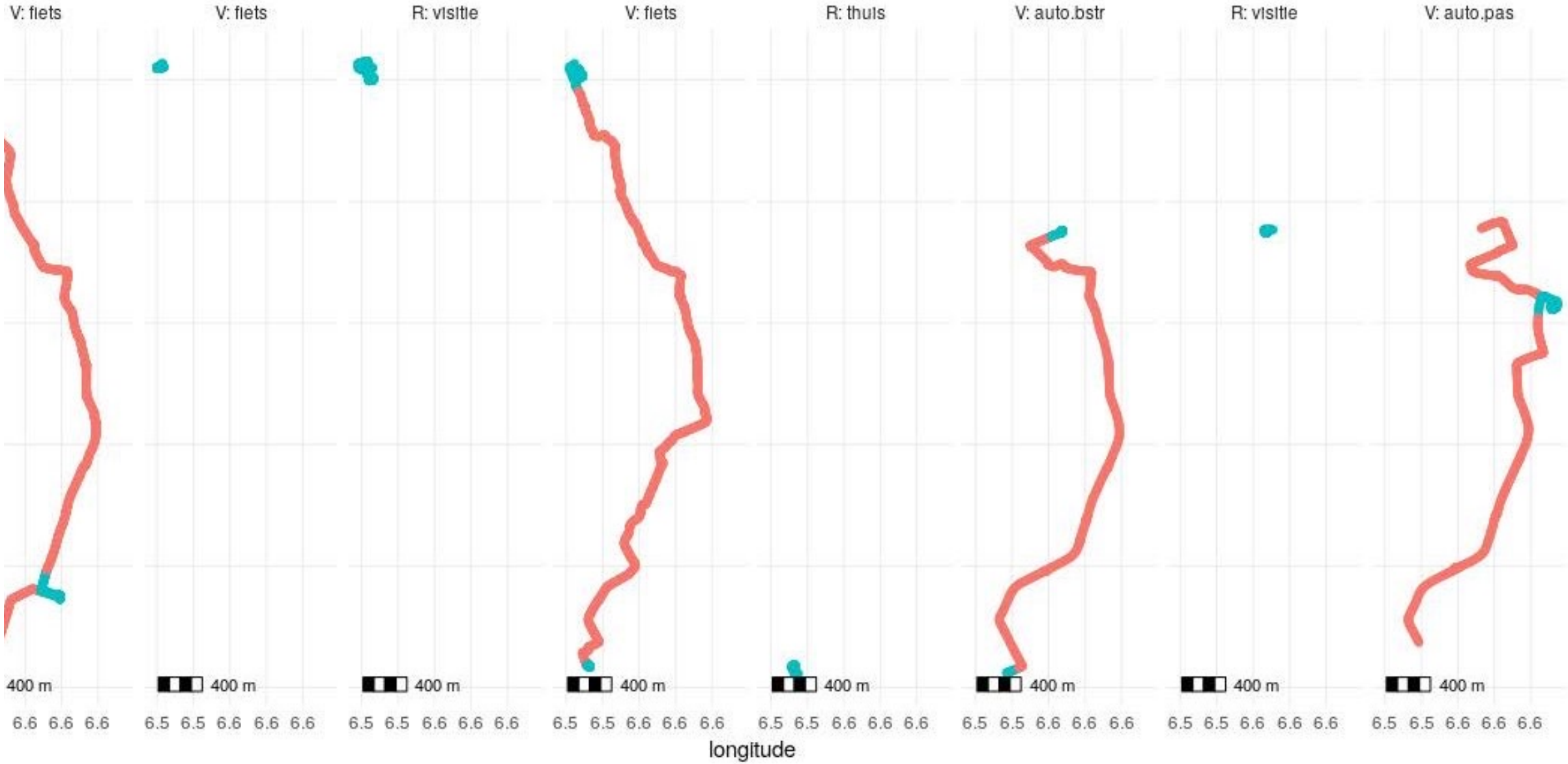
latitude	longitude	timestamp
<num>	<num>	<POSc>
51. 5	5. 8	2022-11-24 13:02:03
51. 7	5 3	2022-11-24 13:02:06
51. 7	5. 3	2022-11-24 13:02:06
51. 1	5. 7	2022-11-24 13:02:
51. 4	5. 0	2022-11-24 13:02:







1. Get location at certain frequency
2. Save the data
3. Filter noise
4. Decide on tracks/staypoints
5. Segment routes
6. User annotations


V: fiets
V: fiets
R: visitie
V: fiets
R: thuis
V: auto.bstr
R: visitie
V: auto.pas
400 m
6.6 6.6 6.6
6.5 6.5 6.6 6.6 6.6
longitude

Dynamic Time Warping Based Imputation

Dynamic Time Warping-Based Imputation

- Aggregate travel characteristic to discrete time
- Set with missingness forms query (q)
- Complete segments form reference sets (ρ)
- Create and interpolate over gaps in ρ, then find best matches with DTW
- Fill gaps in q with created gaps in ρ.