

Data Observer

Florian Zimmermann*, Andreas Filser, Georg-Christoph Haas and Sebastian Bähr

The IAB-SMART-Mobility Module: An Innovative Research Dataset with Mobility Indicators Based on Raw Geodata

<https://doi.org/10.1515/jbnst-2023-0051>

Received July 14, 2023; accepted July 15, 2023

Abstract: The ubiquity of smartphones has enabled the collection of novel data through their built-in sensors, including geolocation data which can be used to understand mobility behavior. In this project, we leveraged longitudinal geolocation data collected from participants in the 2018 German app study IAB-SMART to develop a set of mobility indicators, such as visited unique locations and traveled distance. The indicators can be linked to the Panel Study Labour Market and Social Security (PASS) survey and administrative employment histories. The resulting novel dataset offers a unique opportunity to study the relationship between mobility and labor market outcomes. This article provides an overview of the study, outlines the data preparation process, and the socio-demographic characteristics of the 398 participants of the IAB-SMART-Mobility module. We present the mobility indicators generated from the geolocation data and provide guidance for accessing the Institute for Employment Research's (IAB) data.

Keywords: data manual; geodata; labor market data; PASS data; smartphone data

JEL Classification: C81 methodology for collecting, Estimating, and Organizing Microeconomic data · data access

***Corresponding author: Florian Zimmermann**, Institute for Employment Research (IAB), Regensburger Str. 100, 90478 Nuremberg, Germany, E-mail: florian.zimmermann@iab.de. <https://orcid.org/0000-0001-5645-8342>

Andreas Filser, Georg-Christoph Haas and Sebastian Bähr, Institute for Employment Research, Nuremberg, Germany. <https://orcid.org/0000-0001-5530-3621> (A. Filser). <https://orcid.org/0000-0001-8706-7870> (G.-C. Haas). <https://orcid.org/0000-0003-4234-228X> (S. Bähr)

1 Introduction

Smartphones have become a part of everyday life: people communicate, seek entertainment and information, but also search for jobs or perform work tasks on their mobile devices (Perrin 2017). The access to log files (e.g. log files of calls and short messages (SMS), app use), and the large number of sensors built into smartphones (e.g. GPS, accelerometer) enable the passive collection of data on user behavior and interactions of the users (Malich et al. 2021), such as geolocation data. Passively collecting behavioral data through smartphones provides an unobtrusive opportunity to observe participants free from reactivity biases (Bähr et al. 2022). Thus, geolocations provide researchers with valuable insights into human behavior by investigating questions that cannot be answered using survey data (Bähr et al. 2022; Link et al. 2014; Malich et al. 2021).

The IAB-SMART-Mobility module is a novel dataset offering researchers the opportunity to leverage the potential of data collected by smartphones in combination with a large-scale panel survey (PASS) and administrative employment histories. This module is based on data from the IAB-SMART project. In the IAB-SMART app study, 4,293 participants were randomly recruited from respondents in the “Labour Market and Social Security” (PASS) study who reported using an Android smartphone in wave 11 fielded in 2017 (Kreuter et al. 2020).¹ The field time for the app-based data collection was from January 8, 2018, to August 31, 2018. During this period, the IAB-SMART app utilized passive measurement through smartphone sensors, such as geolocation data (Kreuter et al. 2020). To ensure compliance with European Union General Data Protection Regulation (GDPR), the IAB-SMART app requested participants’ consent to different data collection functions, including *location and network connectivity data* (Bähr et al. 2022). Based on the collected geolocation data by the IAB-SMART app, we generated meaningful mobility indicators in the IAB-SMART-Mobility module.

The IAB-SMART-Mobility module is the first large-scale German dataset with meaningful mobility indicators derived from geolocation data available to the research community. The dataset contains information on participants’ *traveled distance, the share of geodata points at home, number of visited unique locations, and variance in geolocation points*. In total, 618 participants downloaded the app,

¹ Since the app needs extensive access to sensor and mobility data on the smartphone, the app focuses on the Android operating system. This focus on Android is common for smartphone studies, as the iOS operating system does not allow extensive access and the market penetration for other operating systems is not sufficient (Kreuter et al. 2020). For PASS wave 11 respondents, no substantial differences were measured between owners of Android smartphones and other operating systems (Kreuter et al. 2020).

completed the introductory survey, and could be verified as PASS wave 11 respondents. This verification is essential as the IAB-SMART app data can be linked to the rich survey data of PASS (Altschul et al. 2023) and German administrative information on employment histories (PASS-ADIAB, Antoni and Bethmann 2019). In summary, the IAB-SMART-Mobility module and the link to the survey and administrative data provide scholars with a unique opportunity to investigate the effects of labor market outcomes on mobility and vice versa.

In the remainder of this article, we first briefly describe the IAB-SMART project, we lay out our data preparation strategy, and the geolocation data's sample characteristics of the geolocation data. Then, we present the mobility indicators in the IAB-SMART-Mobility module and their construction before concluding with guidance for accessing the data.

2 Data Preparation

Our mobility indicators utilize geolocation data obtained through the IAB-SMART app. The participants' geolocation was collected every 30 min through the app's location and network connectivity functions if participants consented to these functions. The app was designed to gather this information using a two-step process to conserve battery power. First, the app checked if the geolocation was already available in the Android system, for example, because another app recently obtained geolocation data through GPS, the mobile carrier network, or Wi-Fi using Google's Fused Location Provider. Second, if no geolocation data was available in the Android system, the app tried to obtain geolocation information through GPS, the mobile carrier network, or a database of known cell tower locations. This second step did not use Wi-Fi information, as it was built according to a default template that only utilizes native Android functions for compatibility purposes (Bähr et al. 2022). Thus, the IAB-SMART data's available geolocation information was collected via GPS, the mobile carrier network, or Wi-Fi via Google's built-in Fused Location Provider API. These collection methods result in different geolocation accuracies, ranging from high accuracy for GPS to low accuracy for the Fused Location Provider.

We create a dataset of mobility indicators from the raw geolocation data collected by the IAB-SMART app. We initially include all geolocation data sources, and we drop imprecise measurements in the data preparation process. For the data preparation, we partly draw upon the approaches described in recent publications (c.f. Müller et al. 2022; Saeb et al. 2015):

1. We verify participants by comparing their socio-demographic information in the IAB-SMART in-app survey and PASS survey data.
2. We structure the geodata into 30-min intervals.

3. We exclude days without consent to collect geolocation data.
4. We remove geolocation data with errors and inaccurate data.
5. We focus on days with recorded geolocation data for at least 8 h and restrict our sample to participants with a minimum of seven valid days.
6. We construct indicators using the final sample.

2.1 Verifying Participants

Initially, we ensure that smartphone owners installing and using the IAB-SMART app are PASS participants since the IAB-SMART app data can be linked to the PASS data. After installing the IAB-SMART app, participants answered a short survey within the app containing questions about their demography (Kreuter et al. 2020). To verify that the PASS and the IAB-SMART study participants are identical, we compare the date of birth and gender from this IAB-SMART app survey to PASS answers. If at least one variable does not match, we drop the observation. We can verify that 618 persons from the PASS panel installed the app and participated in IAB-SMART.

Table 1 displays the distribution of the first installation and last signal from the app. In sum, most participants installed the app at the beginning of the survey and sent information until around the survey's end: The majority of the 618 verified participants installed the app within a month after the beginning of the field time on January 8, 2018. Specifically, 75 % of the verified participants installed the app by February 10, 2018, i.e. within a month after the start of field time. Turning to the latest signals, 50 % of participants had the app installed until the end of field time on August 30, 2018.

2.2 Structuring Data into 30-min intervals

After restricting our sample to verified participants, we structure the raw geodata to harmonize the time of measurements across participants. The app aimed to collect geolocation information every 30 min, yet the measurements were not taken at precisely the same time for each participant. In order to harmonize the

Table 1: Participants' first installation and last information.

	Min	Q25	Median	Q75	Max	N
First installation	10-Jan-2018	31-Jan-2018	03-Feb-2018	10-Feb-2018	10-Aug-2018	618
Last information	27-Jan-2018	29-Aug-2018	31-Aug-2018	31-Aug-2018	31-Aug-2018	618

Notes: Own calculations using IAB-SMART data.

measurement times, we structure the measurements into 30-min intervals starting at 0:00 am. Based on the 618 participants, we create a raw data file with 48 intervals for each 30-min interval per field day. We then assign the geolocation measurements to the closest interval before the measurement. For example, if a geocode was measured between 9:00 and 9:29, we assign this measurement to the 9:00 measurement interval (Figure 1). Allocating geocodes into intervals allows us to structure the data and ensures that data is comparable between persons.

2.3 Ensuring a Participant’s Consent

In the next step, we focus on geolocation measurements for which participants gave consistently consent daily. Users can grant or deny access to the smartphone *location and network connectivity data*, which is necessary for the geolocation data collection by the SMART app. We remove 51 participants from the sample because they did not consent to collect geodata on a single day.

2.4 Cleaning Geolocation Data

We remove problematic geolocation measurements for the remaining 567 participants who consented to data collection. There are three types of measurement problems in the raw geolocation records. First, some measurements have failed and recorded a longitude and latitude position of 0. Missing geolocation measurements might occur because of a participant’s phone running out of battery, getting switched off, or losing network service (see Bähr et al. 2022). We remove these measurements of 0/0, which reduces the sample to 543 participants with 1,465,541 measurement points.

A second problem is imprecise geolocation records. Geolocation measurements are not precise points but come with a radius of accuracy (e.g. 350 m) describing the area where the geolocation might be. Geolocations were recorded using different

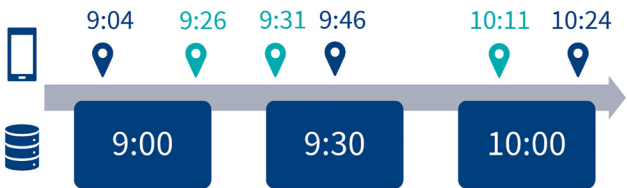


Figure 1: Structuring the data into 30-min-intervals: measurements categorized into.

Table 2: Provider and accuracy of geolocation data.

Provider	Min	Median	75 %	85 %	90 %	Max	<i>N</i>
GPS	0.0	13.0	28.0	45.0	51.0	732.0	223,254
Network	1.0	20.7	34.4	98.4	800.0	10,718,928.0	1,200,289
Fused	2.4	4,898.4	10,014.4	13,804.8	75,130.9	10,718,928.0	46,174
Total	0.0	20.6	36.0	96.1	900.0	10,718,928.0	1,469,717

Notes: Own calculations using IAB-SMART data.

collection providers associated with different levels of measurement accuracy. Table 2 displays the measurement accuracy for all geolocation records by type of measurement provider. GPS localization is substantially more precise than geolocation based on Google’s Fused Location Provider: the median accuracy is 13.0 m for GPS, while measurement using the Fused Location Provider is 4,898.4 m are in line with previous studies (e.g. Zandbergen 2009). Following the literature (e.g. Jongs et al. 2020), we focus on geolocation data with an accuracy of no more than 350 m. Since we restrict our sample to geolocation points with sufficient accuracy, we do not exclude specific providers.

A third measurement problem concerns the collection of multiple geolocation measurements per 30-min interval for approximately 25,000 intervals, which is less than 0.02 % of the intervals. The duplicates were an exact match in approximately 75 % of the cases. In the remaining cases, we keep the measurement with higher accuracy. In sum, this yields a data set of 1,288,050 geolocation measurements from 543 participants that meet the minimum requirements regarding individual measurement quality.

2.5 Ensuring Participants’ Sufficient Observation Periods

In addition to including only high-quality records at the measurement level, we also ensure a minimum number of measurements at the participant level. Specifically, we restrict the data to participants with at least seven ‘valid’ days, i.e. days with at least 8 h with sufficient geolocation records. Following the example provided by Müller et al. 2022, we first check whether at least one geolocation measurement was recorded in a given hour. We restrict the dataset to valid days with at least 8 h and at least one geolocation record. We focus on participants with a minimum of seven valid days. Furthermore, we exclude days when app usage data reveals that participants used geolocation spoofing apps (e.g. Fake-GPS). This step leaves us with 1,218,715 geolocation measurements from 398 participants.

2.6 Summary of Data Preparation and Sample Description

In summary, our data preparation yields a final data set of 398 out of the initial 618 verified participants. On the level of individual geolocation measurements, 1,218,715 out of 1,465,541 usable measurements remain (Table 3). Our final sample comprises 64 % of the verified participants and 83 % of the geolocation measurements. The discrepancy in the number of removed participants and geolocations illustrates that we remove primarily participants with few geolocation measurements. Thus, we only remove participants who had very few geolocation measurements, for whom mobility indicators are of little use due to gaps in the data. See Appendix A for a detailed overview of the data collection (following Müller et al. 2022).

Participants in our final sample differ in socio-demographic characteristics from the general population. For instance, our sample overrepresents 17-39-year-old participants compared to the German population (Destatis 2023). 34 % of participants in our sample are located in new federal states, i.e. eastern Germany, which is higher than the share of Germans living in eastern Germany (20 %) (Kreuter et al. 2020). The overrepresentation of younger participants and participants in eastern Germany is similar to the PASS (Kreuter et al. 2020). Unemployed and inactive persons are overrepresented in our sample (36 %) in comparison to Germany in 2017 (6 %) (Kreuter et al. 2020), which can be attributed to PASS’s oversampling of unemployed persons. In summary, the sample similarly deviates from the German population, given that IAB-SMART participants are randomly drawn from the PASS, yet the deviations for IAB-SMART are larger. Scholars need to take these deviations into account when analyzing the IAB-SMART-Mobility module.

3 Indicators

We generate four mobility and three data-quality indicators using the prepared geolocation data. For each participant, the *mobility indicators* are aggregated at the

Table 3: Number of persons and geolocation at each stage of data preparation.

Step	Persons	% persons	Geodata	% geodata
Installed and verified	618	100 %		
Geodata consent	567	92 %		
Valid geolocation measurements	543	88 %	1,465,541	100 %
Accurate geolocation measurements (<350 m)	543	88 %	1,288,050	88 %
At least seven valid days	398	64 %	1,218,715	83 %

Notes: Own calculations using IAB-SMART data.

weekday, i.e. Monday to Friday, and weekend, i.e. Saturday and Sunday, levels to represent average routines on typical workdays and non-workdays. Furthermore, for each participant, we aggregate the total data collected over the entire study period to provide a general level of mobility. The *data-quality indicators* provide users with a measure of data quality, such as the number of geolocation points (see Table 5). The *data-quality indicators* are calculated separately for the total data collection period and weekdays and weekends.

Table 4 displays summary statistics of the four *mobility indicators*. For the summary statistics of the total data collection period, we compute summary statistics on the participant level. For the weekday and weekend summary statistics, we first compute the participants' median on weekdays and weekends. Second, we use the mean values to compute the summary statistics in Table 4.

3.1 Visited Unique Locations

The first indicator captures the number of unique locations visited per participant. Locations are identified following the clustering of individual geolocation points using the density-based spatial clustering of applications with noise method (DBSCAN) (Hahsler et al. 2019; Jongs et al. 2020; Müller et al. 2022). For the DBSCAN clustering, we chose a range of 600 m, similar to Jongs et al. (2020), since we also follow the accuracy restriction of 350 m recommended by Jongs et al. (2020). The clustering range needs to be higher than the accuracy to ensure that clusters are not biased by inaccurate measurements. Second, we chose two as the minimum geolocations per cluster. A location represents a place participants visit and spend time.

Table 4: Summary statistics of mobility indicators derived from geolocation data.

Variable	Total			Weekday (median day)			Weekend (median day)		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Visited unique locations	25.4	25.6	398	1.8	0.9	398	1.5	0.6	398
Traveled distance (in km)	7,987.2	39,698.9	398	40.0	419.2	398	46.1	629.2	398
Share measurements at home during daytime	76.5	18.1	375 ^a	82.2	20.8	375 ^a	91.8	14.5	372 ^a
The variance of locations (log)	−2.7	4.9	398	−9.4	5.0	398	−11.5	5.2	397 ^b

Notes: ^aWe cannot identify the home of some participants due to a lack of sufficient observations at night. ^bOne participant has a variance of locations of 0 at the weekend, resulting in a missing due to the logarithmic transformation. Own calculations using the IAB-SMART-Mobility module.

Since geolocation was measured in 30-min intervals, two geolocation measurements from the same cluster correspond to a stay of 30–60 min or at least two visits to that geolocation at different times. Participants visited 25.4 unique locations within the study period (Table 4). The average number of visited places is slightly higher on weekdays than on weekends (Table 4).

3.2 Traveled Distance

We calculate the distances between two consecutive points to measure the traveled distance. Given that we are using an extensive study region without any limits (participants cross states and even travel abroad), we use the *distHaversine* function from the *geosphere* package (Hijmans et al. 2017) to account for the curvature of the Earth when calculating distances (Müller et al. 2022). The Haversine distance calculates the shortest distance between two points (i.e. the ‘great circle distance’ or ‘as the crow flies’), assuming a spherical earth, ignoring ellipsoidal effects. Based on the clustering, distances are set to zero for consecutively measured points in the same visited unique location. Consequently, the total distances capture the movement between unique locations as well as solitary geolocations. We aggregate the traveled distances between two locations to calculate the traveled distance on a median weekday and weekend, as well as the total traveled distance. Data users can calculate the mean distance traveled per day from the total distance and number of days with measurement.

3.3 Share Measurements at Home during Daytime

We identify homes following the strategy employed by Müller et al. (2022), who use two conditions: First, the home is the location with the most geodata points at night, i.e. between 0 am and 6 am. Second, to qualify as an identified home cluster, the respective location cluster must be the modal geolocation cluster for at least three nights (Müller et al. 2022, p. 5). After identifying the home cluster, we calculate the share of geolocations from the home cluster during the daytime, between 6 am and 0 am. Since geolocations at night were used to classify homes, we exclude geolocations at night for the share of geodata at home classification, that is, geolocations between 0 am and 6 am. For 23 participants, no home could be identified due to a lack of measurements at night. On average, participants are located in 82.2 % of their geodata points at home during a weekday and 91.8 % during the weekend (Table 4).

3.4 Location Variance

Location Variance = $\log (\sigma_{\text{lat}}^2 + \sigma_{\text{long}}^2)$

The location variance is calculated using the variance of a person’s latitudinal and longitudinal geolocation records. This variance represents the variability in a participant’s geolocation (Saeb et al. 2015). A higher variance indicates that the measured geodata records are further apart, indicating a higher radius of activities. Since the variance of geolocation points is usually below 1, the log of location variance is negative. A higher value, i.e. closer to 0, indicates a higher variance in geodata records. Similar to the previous indicators, we calculate the location variance for the entire study period and separately for weekdays and weekends.

3.5 Data-Quality Indicators

To provide data users with measures of the underlying data quality for each participant, we calculate three *data-quality indicators* for the total observation period, weekdays, and the weekend: First, the *number of geolocation points* summarizes the total number of geolocations that we use for calculating the mobility indicators for each participant. On average, 3,070 data points were used for each participant.

Second, we provide the number of valid days per person. As described in the section *Ensuring Participants’ Sufficient Observation Periods*, we consider days valid if at least 8 h of geolocation data are available. On average, participants provided sufficient geolocation data for 106.5 days (Table 5).

Third, we provide the average share of geodata records collected per day. Here, we calculate the total collected records and divide by 48, the theoretical maximum of collected points, to calculate this indicator. On average, 56.3 percent of the theoretical maximum is recorded (Table 5). Users can use these quality indicators to restrict

Table 5: Summary statistics of data quality indicators.

Variable	Total			Weekday			Weekend		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Number of geolocation points	3070.0	2351.1	398	2209.6	1699.1	398	860.4	661.3	398
Number of valid days	106.5	69.5	398	76.3	50.0	398	30.2	19.8	398
Share of valid geodata	0.6	0.2	398	0.6	0.2	398	0.6	0.2	398

Notes: Own calculations using the IAB-SMART-Mobility module.

the sample to their research needs. For example, the total distance and number of locations correlate with the number of days and the total number of measurement points.

4 Conclusion and Guidance on Data Access

In this article, we described how we generated mobility indicators based on geolocation data for the IAB-SMART-Mobility module. We described how the geolocation data was collected, which preprocessing step we took, and how we built meaningful mobility indicators. Due to its unique link to the PASS panel survey data and administrative employment records, this data allows for analyzing the association between labor market outcomes and activities.

When using the IAB-SMART-Mobility Module, researchers should keep in mind the scope and informative value of the data. As such, aggregating raw geolocation data to meaningful indicators involves many decisions on data preprocessing steps. Each decision may substantially affect the relationship between input and outcome variables in the models. While building our indicators, we aggregated the information at the weekday and weekend levels to ensure participants' confidentiality while still reflecting participants' activities in our indicators. If researchers require a more fine-grained aggregation of indicators to investigate a labor market research question, we invite researchers to contact the IAB-SMART team² for potential cooperation.

Due to IAB-SMART's novel data types (i.e. sensor and smartphone usage data) and its unique embedment in the PASS probability sample, initial work focused on methodological questions. Studies based on smartphone information potentially rely on selective data. Sampling android smartphone owners from within the PASS population, allowed Keusch et al. (2023) to assess this coverage error. Studies focusing on the geodata, include Haas et al. (2020) who share lessons from using geofences as survey triggers and Bähr et al. (2022) who analyze the conditions for observing valid geoinformation in the presence of sources of missingness inside and outside the smartphone. This ample methodological work lays the foundation for the novel IAB-SMART-Mobility module.

The indicators provided by the IAB-SMART-Mobility module can be used to contribute to multiple strands of literature. First, literature on time use (e.g. Krueger and Mueller 2012a) indicates that unemployed individuals use time differently than employed individuals and that these patterns especially differ on weekdays and the

2 If you are interested, please contact: E-mail: sebastian.baehr@iab.de, georg-christoph.haas@iab.de or mark.trappmann@iab.de.

weekend (Krueger and Mueller 2012b). These findings using time diary data could be complemented using passively collected smartphone data as provided by IAB-SMART-Mobility's the mobility indicators. Second, research on mental health has highlighted the importance of mobility for mental well-being (Jongs et al. 2020; Saeb et al. 2015). For instance, smartphone-location-based indicators similar to those provided by IAB-SMART-Mobility have been shown to correlate with the severity of depressive symptoms (Saeb et al. 2015). Combining the IAB-SMART-Mobility indicators with information on employment and mental health from the PASS survey opens up avenues for future research at the intersection of psychology and labour market research.

The IAB-SMART-Mobility module is the first dataset that offers the international scientific community access to meaningful mobility indicators in Germany. The IAB-SMART-Mobility module can be accessed with the PASS scientific use file (SUF) via the FDZ (<https://fdz.iab.de/en/data-access/scientific-use-files/>). See Appendix B for a brief Stata code example to merge the module with the PASS dataset.

Due to its lower degree of anonymization, linkage of the IAB-SMART-Mobility module and PASS to the administrative employment histories, the PASS-ADIAB (Antoni and Bethmann 2019) is only available for on-site use or remote execution using JoSuA (Eberle et al. 2017). Access to the data is available for labor market research and is free of charge. Methodological research is not possible. Users have to enter into a data use agreement with the FDZ and fulfill further requirements depending on the access mode. Additional information can be found on the FDZ homepage (<https://fdz.iab.de/>).

Acknowledgments: We thank BERD@NFDI for funding this project. We also thank Dana Müller and Mark Trappmann for their helpful comments and Florian Keusch and Frauke Kreuter for sharing their expertise on the IAB-SMART data.

Appendix A: Information about how the GPS data were collected (following Müller et al. 2022)

- Number of participants in the study (recruited, dropped out, excluded, and retained in the final sample used for analyses) **4,293 total, 618 installed and verified (14.6 %), 398 (9.3 %) after all restrictions**
- Geographical area covered (by design or characteristics of the data set): **No restriction**
- Sampling frequency (how often latitude and longitude coordinates were collected): **30 min**

- Type of sampling strategy: **Periodically**
- Accuracy of GPS points: **for raw data see Table 2; for final sample Accuracy < 350 m (Hahsler et al. 2019; Jongs et al. 2020)**
- Duration of data collection: **January 8 2018 to August 31 2018**
- Operating system used to collect the data: **Android Version 4.0 or higher**
- App/data collection setup used: **IAB-SMART App** (see Kreuter et al. 2020)
- Number of GPS and network records collected: **In total: 1,465,541; final sample: 1,218,715**

Appendix B: Code example Stata 17: Merging IAB-SMART-Mobility and PASS SUF

```
//Defines path for the datasets
global orig_pass //Path for PASS_0621_v2
global orig_smart //Path for IAB_SMART

//Merge the data and keep only wave 11 from PASS
use "${orig_smart}/IAB-SMART-Mobilityindicators.dta", clear
merge 1:m pnr using "${orig_pass}/PENDDAT.dta", keep (match master)
keep if welle == 11 //Keep only wave 11 (from 2017)
```

References

- Altschul, S., Bähr, S., Beste, J., Collischon, M., Coban, M., Dummert, S., Frodermann, C., Gleiser, P., Gundert, S., Küfner, B., et al. (2023). *Panel Arbeitsmarkt und soziale Sicherung (PASS) – Version 0621 v2*. Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nuremberg.
- Antoni, M. and Bethmann, A. (2019). PASS-ADIAB–linked survey and administrative data for research on unemployment and poverty. *Jahrb. Natl. Stat.* 239: 747–756.
- Bähr, S., Haas, G.C., Keusch, F., Kreuter, F., and Trappmann, M. (2022). Missing data and other measurement quality issues in mobile geolocation sensor data. *Soc. Sci. Comput. Rev.* 40: 212–235.
- Destatis (2023). *Bevölkerung nach Altersgruppen 2011 bis 2021 in Prozent*, Available at: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Tabellen/bevoelkerung-altersgruppen-deutschland.html> (Accessed 10 March 2023).
- Eberle, J., Müller, D., and Heining, J. (2017). A modern job submission application to access IAB’s confidential administrative and survey research data. In: *FDZ methodenreport, 1, 2017*. Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB), Nuremberg.
- Haas, G.C., Trappmann, M., Keusch, F., Bähr, S., and Kreuter, F. (2020). Using geofences to collect survey data: lessons learned from the IAB-SMART study. *Survey Methods: Insights from the Field*, 1–12.

- Hahsler, M., Piekenbrock, M., and Doran, D. (2019). Dbscan: fast density-based clustering with R. *J. Stat. Software* 91: 1–30.
- Hijmans, R.J., Karney, C., Williams, E., Vennes, C., and Hijmans, M.R.J. (2017). Package ‘geosphere’. *Spherical Trigonometry* 1: 1–45.
- Jongs, N., Jagesar, R., van Haren, N.E.M., Penninx, B.W.J.H., Reus, L., Visser, P.J., van der Wee, N.J.A., Koning, I.M., Arango, C., Sommer, I.E.C., et al. (2020). A framework for assessing neuropsychiatric phenotypes by using smartphone-based location data. *Transl. Psychiatry* 10: 211.
- Keusch, F., Bähr, S., Haas, G.C., Kreuter, F., and Trappmann, M. (2023). Coverage error in data collection combining mobile surveys with passive measurement using apps: data from a German national survey. *Socio. Methods Res.* 52: 841–878.
- Kreuter, F., Haas, G.C., Keusch, F., Bähr, S., and Trappmann, M. (2020). Collecting survey and smartphone sensor data with an app: opportunities and challenges around privacy and informed consent. *Soc. Sci. Comput. Rev.* 38: 533–549.
- Krueger, A.B. and Mueller, A.I. (2012a). Time use, emotional well-being, and unemployment: evidence from longitudinal data. *Am. Econ. Rev.* 102: 594–599.
- Krueger, A.B. and Mueller, A.I. (2012b). The lot of the unemployed: a time use perspective. *J. Eur. Econ. Assoc.* 10: 765–794.
- Link, M., Murphy, J., Schober, M.F., Buskirk, T.D., Hunter Childs, J., and Langer Tesfaye, C. (2014). Mobile technologies for conducting, augmenting and potentially replacing surveys: executive summary of the AAPOR task force on emerging technologies in public opinion research. *Publ. Opin. Q.* 78: 779–787.
- Malich, S., Keusch, F., Bähr, S., Haas, G.C., Kreuter, F., and Trappmann, M. (2021). Mobile Datenerhebung in einem Panel – die IAB-SMART Studie. In: Wolbring, T., Leitgöb, H., and Faulbaum, F. (Eds.). *Sozialwissenschaftliche Datenerhebung im digitalen zeitalter*. Springer Vs, Wiesbaden, pp. 45–69.
- Müller, S., Bayer, J.B., Ross, M.Q., Mount, J., Stachl, C., Harari, G.M., Chang, Y.J., and Le, H.T.K. (2022). Analyzing GPS data for psychological research: a tutorial. *Adv. Meth Pract. Psychol Sci.* 5: 1–11.
- Perrin, A. (2017). *10 Facts about smartphones*. Pew Research Center, Available at: <https://www.pewresearch.org/fact-tank/2017/06/28/10-facts-about-smartphones/> (Accessed 6 February 2023).
- Saeb, S., Zhang, M., Karr, C.J., Schueller, S.M., Corden, M.E., Kording, K.P., and Mohr, D.C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *J. Med. Internet Res.* 17: e4273.
- Zandbergen, P.A. (2009). Geocoding quality and implications for spatial analysis. *Geogr. Compass* 3: 647–680.