Data Descriptor for Nature Scientific Data

Carmen Cabrera-Arnau Francisco Rowe
Miguel González-Leonardo Andrea Nasuto Ruth Neville

2024-12-01

COVID-19 triggered a reduction in the frequency and extent of people's movement. Existing evidence suggests that while the impact of the pandemic on mobility was widespread, the extent of this impact was unequally felt across socioeconomic groups in the early stages of the pandemic. Here, we find that the most deprived locations have experienced a more accelerated recovery towards pre-pandemic levels of mobility in the long term. Furthermore, the socioeconomic disparities in the patterns of mobility triggered by the first outbreak of COVID-19, have persisted as of April 2023. These findings are based on the analysis of time-series mobility data corresponding to X urban areas from Latin American countries collected from Meta-Facebook users upon their consent. Our research highlights the importance of timely mobility data with high spatiotemporal resolution to understand the long-term effects of the pandemic and to inform equitable policy responses that address societal challenges in urban areas.

Main

@Francisco to rewrite this section

Spatial human mobility is key to creating sustainable, livable and inclusive cities. At the societal level, spatial mobility enables the transfer of knowledge, skills and labour to places they are needed (Ackers 2005). Spatial mobility also shapes service and transport demand across urban spaces (Chen et al. 2016), and enables the monitoring and control of transmissible diseases (Belik, Geisel, and Brockmann 2011). At the individual level, mobility enables people to access and achieve opportunities and aspirations in space (Klugman 2009). Understanding spatial human mobility is thus important to supporting appropriate policy responses to address societal challenges relating to carbon emissions, urban planning, service delivery, public health, disaster management and transport (Barbosa et al. 2018; Chinazzi et al. 2020).

The COVID-19 pandemic resulted in a notable decrease in mobility, particularly in cities (Nouvellet et al. 2021). Coupled with fears of contagion in crowded public spaces, nonpharmaceutical interventions to contain the spread of COVID-19 prompted this decrease in the overall levels of mobility (Nouvellet et al. 2021; Rowe, González-Leonardo, and Champion 2023). Especially during lockdowns, mobility recorded reductions in the frequency, distance and time of trips in cities across the globe (Abdullah et al. 2020; Bonaccorsi et al. 2020; Abu-Rayash and Dincer 2020; Lee et al. 2023). Higher engagement with remote working, online schooling and shopping activity reduced the need to travel for work, education, shopping and leisure, hence giving rise to more geographically localised mobility patterns (Borkowski, Jażdżewska-Gutta, and Szmelter-Jarosz 2021).

However, reductions in mobility levels were highly unequal reflecting existing socioeconomic inequalities in our societies (Chang et al. 2020). In most countries, affluent individuals tended to record the greatest drops in mobility levels as they are predominantly employed in knowledge-intensive jobs which can be done fully or partly remotely (Fraiberger et al. 2020; Bonaccorsi et al. 2020; Weill et al. 2020; Dueñas, Campi, and Olmos 2021; Santana et al. 2023). During the COVID-19 pandemic, the adoption of remote work reduced the need of commuting for knowledge-intensive, non-public facing jobs (Florida, Rodríguez-Pose, and Storper 2021). At the same time, individuals from less privileged socioeconomic backgrounds displayed less pronounced declines mirroring the nature of their work requiring public-facing, face-to-face interaction, and thus a requirement for daily work commutes (Dueñas, Campi, and Olmos 2021; Santana et al. 2023).

Thus, while a growing body of empirical evidence has contributed to advancing our understanding of the impacts of the COVID-19 pandemic on spatial mobility within cities, existing research has focused on more developed countries and the immediate effects of the pandemic during 2020. Less is known about the longer term patterns of resilience in urban mobility in less developed countries extending beyond this period (Rowe et al. 2023). Urban spaces have changed considerably since then, from going through waves of high COVID-19 fatality, infections, school and business closures to the removal of all COVID-19 restrictions as the UN World Health Organization (WHO) declared an end to the pandemic as a public health emergency; yet, different configurations of hybrid working have remained in the norm across many sector of the economy (Barrero, Bloom, and Davis 2021; Aksoy et al. 2022). Thus, assessing the extent to which the level of mobility has returned back to the pre-pandemic baseline level across socioeconomic groups is important to understand the potentially unequal long-term impacts of hybrid working.

A key barrier to monitor changes in geographic mobility patterns in less developed countries during and post the COVID-19 pandemic has been the lack of suitable data (Rowe et al. 2023). Traditionally census and survey data have been employed to analyse human mobility patterns in these countries (Green, Pollock, and Rowe 2021). Yet, these data streams are not frequently updated and suffer from slow releases, with census data for example being collected over intervals of ten years in most countries (Bell et al. 2014). Traditional data streams thus lack the temporal granularity to analyse population movements over short-time periods and to offer an up-to-date representation of the urban mobility system (Rowe 2023b). Data resulting from social interactions on digital platforms have emerged as an unique source of information to deliver this representation and capture human population movement in less developed countries at scale (Rowe 2023b). Particularly location data from mobile phone applications have become a prominent source to sense patterns of human mobility at higher geographical and temporal resolution in real time (Calafiore et al. 2023).

Drawing on a dataset of 213 million observations from Meta-Facebook users' mobile location data, we aim to assess socioeconomic differences in the extent and persistence of decline in urban mobility in Argentina, Chile, Colombia and Mexico during and after the COVID-19 pandemic from March 2020 to March 2023. We use Meta-Facebook data to measure origin-destination flows from March 2020 to May 2022, and Meta Prophet time-series forecasting machine learning algorithm (Taylor and Letham 2017) to predict origin-destination flows from June 2022 to March 2023. We use Functional Urban Areas (FUAs) boundaries from the Global Human Settlement Layer (GHSL), developed by the European Comission's Joint Research Centre (Schiavina M. 2019) to define urban areas; and the Global Gridded Relative Deprivation Index (GRDI) developed by NASA's Socioeconomic Data and Applications Centre (Columbia University 2022) from sociodemographic and satellite data inputs. Building on existing evidence (e.g. Rowe et al. 2022; Wang et al. 2022), we hypothesised that (1) urban mobility has recovered returning to the pre-pandemic baseline level of movement as nonpharmaceuthical restrictions were lifted; and, that (2) socioeconomic differences in urban mobility have endured the pandemic reflecting deep societal inequalities as knowledge-intensive businesses adopt hybrid working.

Latin America provides an ideal test-bed for testing these hypotheses because of its exceptionally high levels of inequalities (De Ferranti 2004; Carranza, De Rosa, and Flores 2023) and urbanisation (United Nations and Affairs. 2023). Half of the 20 most unequal countries in the planet are in this region. The average income Gini index of the region is 4 percentage points higher than that of Africa and 11 higher than China (Milanovic 2016), and cities display some of the starkest inequalities (Habitat 2022). Currently, over 80% of the population in Latin America live in urban areas. By 2050, this share is predicted to reach 89%, with the largest share concentrating in a few megacities (Habitat 2022). Developing an understanding of human mobility in Latin America is thus important to support sustainable and inclusive spaces (Habitat 2022).

Results

We focus on capturing how mobility patterns for different population groups have been impacted by COVID-19 in Latin America. We consider populations groups by classifying the spatial units of analysis into categories according to their population density and their relative level of socioeconomic deprivation. More details for the classification method are provided in the Methods section. The two criteria for classification are chosen due to their relevance for

`@Francisco, can you help here?`. The analysis of mobility is done using Facebook movement data that has been processed according to our methodology.

Affluent and densely populated areas saw steepest decline in daily mobility, but all population groups progress towards baseline levels

We first focus on analysing the variability in COVID-19's impact on daily mobility levels across different population groups. Figure 1 shows changes in the intensity of movement measured as the percentage change in the number of inflows relative to pre-pandemic baseline levels, at two key points during the period of analysis, May 2020 and March 2022. These two months are the most widely spaced months with complete data available for all countries within the period covered by the dataset. May 2020 represents the initial phase following the WHO's declaration of COVID-19 as a global pandemic on March 11, 2020, marked by stringent measures. March 2022 represents a later phase, roughly six months after most restrictions had been lifted in the countries included in the analysis. Figure 1 presents boxplots depicting the distribution of percentage change in movement counts across areas grouped by population density and deprivation index. The boxplots show mobility changes for each category in May 2020 and March 2022, with baseline levels marked by the dotted line at $y = 0$. Values above this line indicate increased mobility from pre-pandemic levels, while values below show a reduction. These changes show remarkably consistent patterns across countries for both types of classification.

Focusing first on the patterns by population density category, we generally observe a decrease in the number of inflows for all countries and most categories of analysis in the early pandemic days. The extent of this decrease tends to be larger for high-density areas. Although the majority of areas saw a decline in mobility during May 2020, some locations actually recorded higher movement levels compared to the pre-pandemic period, particularly in regions classified as low-density. In fact, some of these areas saw their mobility levels double, as indicated by outliers with percentage change exceeding 100%. However, this does not imply that these areas correspond to large population numbers, as low-density regions typically contain very few residents, meaning that minor fluctuations in population can lead to significant percentage change. Nonetheless, these relative changes illustrate the extent of the impacts experienced by these sparsely populated regions. In March 2022, nearly two years later, mobility levels have generally rebounded closer to pre-pandemic baseline levels across different population density categories. Notably, mobility levels in areas characterised by higher population density still remain below the baseline. Low-density areas show even greater variability than during the early days of the pandemic, with mobility levels often exceeding the baseline. These high variability in sparsely populated regions may reflect the ongoing effects of the pandemic on mobility, but they could also be influenced by seasonal variation (e.g. March compared to May). Overall, the observed patterns by population density category are consistent with those reported in (Rowe et al. 2024) for Argentina and Chile, where data from the same source corresponding to those two countries was analysed without applying the data processing methodology proposed in this study.

Moving to the patterns by relative deprivation index category, we observe tendencies that mirror the patterns by population density category. Firstly, in March 2020, there is an overall decrease in the number of inflows for all countries and most categories of analysis. Exceptions to this general tendency notably occur in the most deprived (less affluent) areas, especially in Colombia, where mobility levels in some regions are actually above the baseline. This decrease displays a gradient across categories of analysis whereby the more affluent an area is, the higher the losses in the number of inflows. After almost two years, we find evidence for a general recovery of mobility closer to baseline levels across groups regardless of how deprived they are. However, disparities between groups are still visible, indicating that the pandemic has exacerbated behavioral differences among groups with varying socioeconomic statuses.

@Miguel, can you elaborate the following in the discussion? - The pattern that we observe in this section can be attributed to... THIS GOES TO THE DISCUSSION

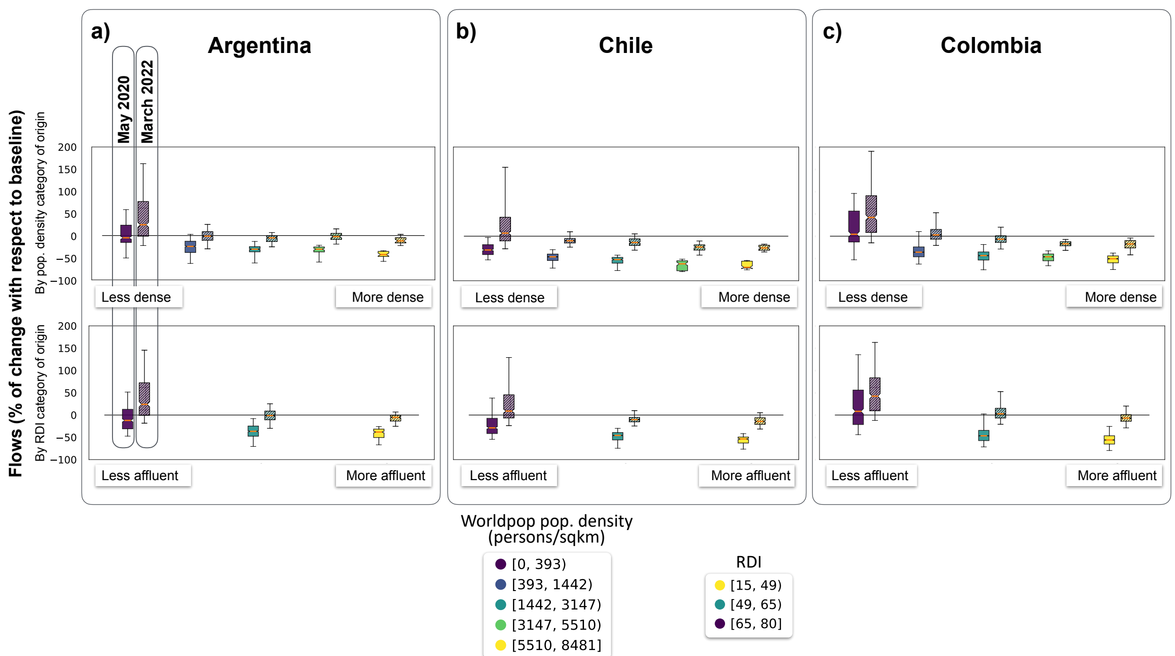

Figure 1: Changes in the intensity of flows as the percentage of change with respect to the pre-pandemic baseline levels, for each country and category of analysis at the origin location of each flow.

Initial drop in mobility drives intergroup variation in mobility levels, as recovery pace is uniform across population groups

Next we focus on modelling intergroup differences in mobility trends over time, highlighting the variability in the pace of recovery toward baseline mobility levels across population groups. Figure 2 and Figure 3 show, for each country, the evolution over time in the intensity of movement, grouped into population density and relative deprivation index categories respectively. Time is represented on the x-axis, spanning from mid-April 2020 to mid-April 2022, while the intensity of movement is depicted on the y-axis, measured as the percentage change in the number of flows relative to pre-pandemic baseline levels. The background of each plot in Figures Figure 2 and Figure 3 is colored according to the stringency index, which quantifies the level of non-pharmaceutical interventions implemented at each point in time to mitigate the spread of COVID-19, including measures such as social distancing and lockdowns. This data provides context for the observed trends, reflecting the mobility restrictions imposed by each country during the specified periods.

@Francisco and @Miguel to check that what I wrote below makes sense

We highlight four key observations from a visual inspection of the evolution patterns by population density category, as shown in Figure 2. First, there is a general upward trend in mobility across all population density categories for all countries, reflecting a recovery of mobility levels following the early days of COVID-19. Second, more densely populated areas consistently record lower levels of mobility throughout the study period. Third, there exist some variations in the patterns by country. Chile experiences greater overall losses in mobility compared to Argentina, while certain areas in Argentina reach levels comparable to the pre-pandemic baseline. In contrast, Colombia displays more varied trends across population density categories, with above-baseline levels of mobility in low-density areas, and significantly larger losses in high-density areas. Fourth, the temporal patterns display fluctuations from the general trend. These fluctuations are likely driven by seasonality effects and by the changing levels of stringency of COVID-19 interventions.

Visual inspection of Figure 3 reveals the following four patterns by relative deprivation index category. First, we observe a general upward trend in mobility across all relative deprivation index categories for all countries. Second, less deprived areas consistently record lower levels of mobility throughout the study period. In contrast, more deprived areas exhibit higher levels of mobility. Third, there is variability by country. Colombia stands out for recording mobility levels in the most deprived areas that are well above the baseline. These big relative changes for the most deprived category do not necessarily correspond to large population numbers since for example, more deprived regions in Colombia are often low-density areas that typically contain small populations. As a result, minor fluctuations in mobility can lead to substantial percentage changes. Nevertheless, the positive percentage change observed in more deprived areas highlight the variability of the impact across different population groups based on their socioeconomic status. Fourth, like in Figure 2, the temporal patterns display fluctuations from

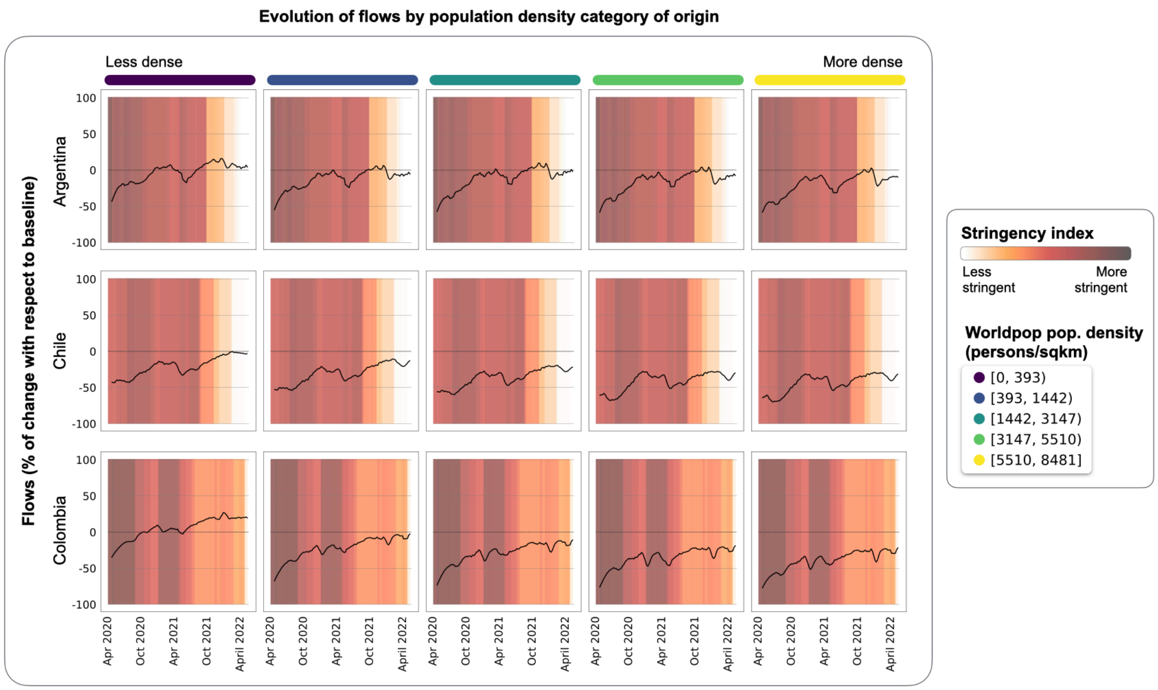

Figure 2: Evolution of number of flows by population density category of origin for Argentina, Chile and Colombia, as percentage change with respect to pre-pandemic baseline levels.

the general trend which are likely driven by seasonality effects and by the changing levels of COVID-19 stringency index.

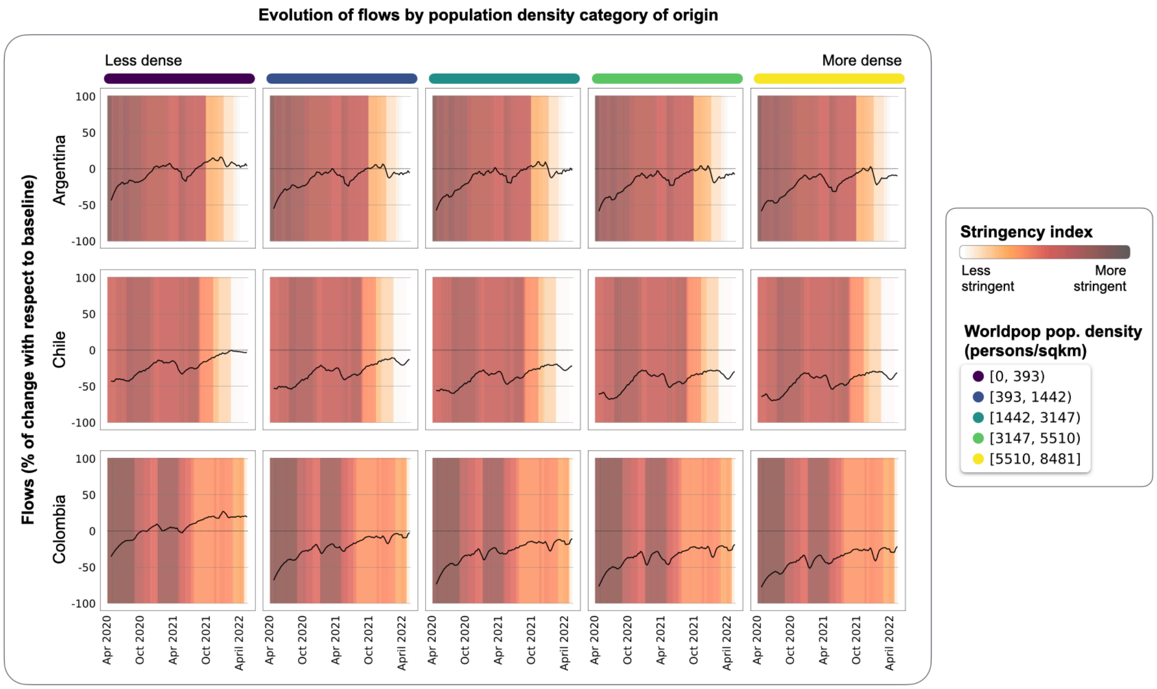

Figure 3: Evolution of number of flows by population density category of origin, as percentage change with respect to pre-pandemic baseline levels.

Next, we quantitatively model intergroup differences in mobility trends over Figure 2 and Figure 3. To achieve this, we perform a time series decomposition into three components: trend, seasonal, and residual. Our analysis focuses on the trend component, which we model using five specifications within a linear mixed-effects modeling framework, all yielding consistent results. All five models include time as a fixed effect and estimate an intercept and a slope associated with this variable. The intercept is interpreted as the magnitude of the initial drop in mobility, while the slope represents the rate at which mobility levels recover over time. Models 3, 4, and 5 incorporate random effects to capture variation in the intercept, slope, or both, based on the category of the origin of the flows. Detailed descriptions of all model specifications are provided in Methodology Section X.

Figures Figure 4 and Figure 5 present, for each country, the random variations in intercept and slope estimated by Models 3, 4, and 5, providing insights into the drivers of intergroup variation in the recovery towards baseline mobility levels. Focusing first on Figure 4, we observe that... *@Francisco to write this interpretation*

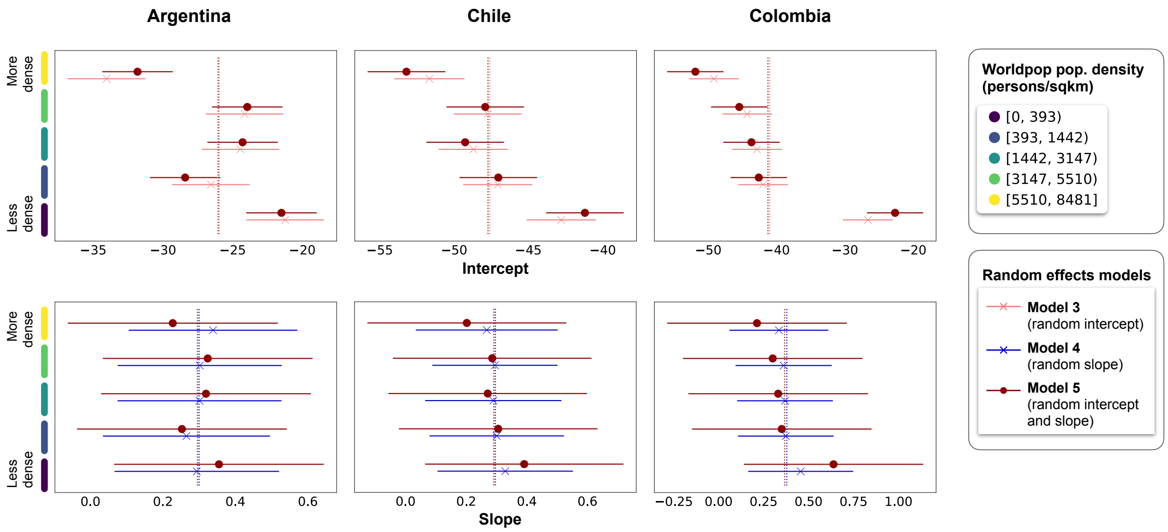

Figure 4: Figure random effects density

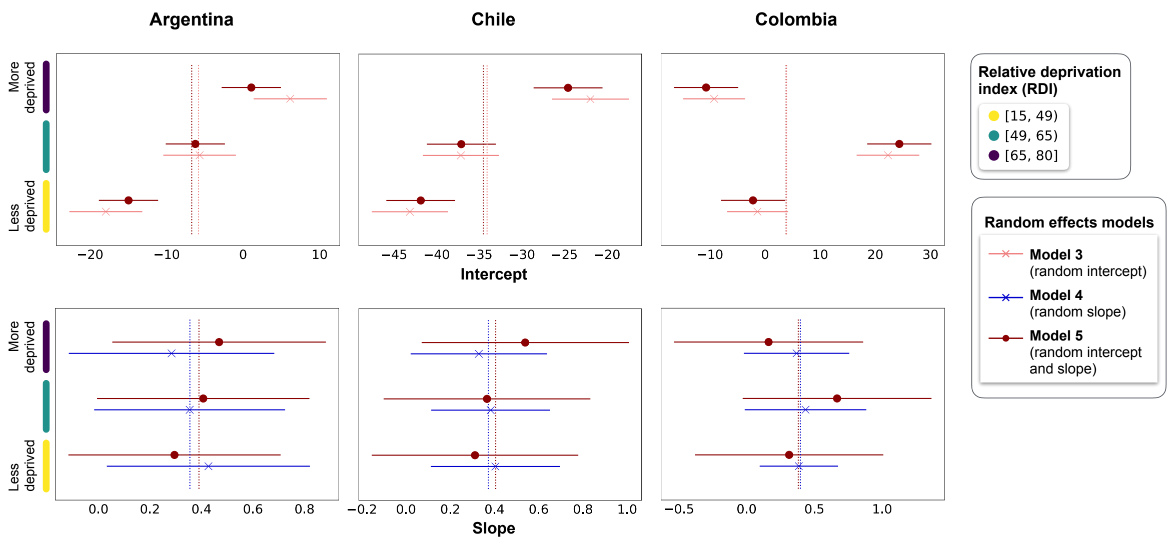

Figure 5: Figure random effects RDI

Focus on cities - *@Francisco @Miguel, we need a better title here - any ideas?*

Turning our focus to cities, we present a city-centric analysis of the evolution of movements. Specifically, we examine movements between urban centers and the rest of spatial units within each country, categorising these spatial units by population density to account for variability across the urban hierarchy. Urban centers are defined as the most densely populated spatial unit within each Functional Urban Area (FUA). The boundaries of these Functional Urban Areas are derived from the European Commission's Human Settlement Layer (GHSL) project (CEU. JRC. 2019) .

For each country, we analyze inflows to all FUA centers, considering two types of movements: those originating from any spatial unit (first column of Figure 6) and those from spatial units within a 100 km radius of the FUA centers (second column of Figure 6). Additionally, focusing only on urban centres corresponding to the capital FUAs -Buenos Aires, Santiago and Bogotá- we also look at inflows from any spatial unit and from those situated within a 100 km radius (columns third and fourth columns in Figure 6). The analysis is similarly repeated for outflows from urban centers to other spatial units.

@Miguel to write this section. What can we learn about the figure? The target journal is nature cities, so we need to talk about cities...

Discussion

@Miguel to revisit this section - the text here was written a long time ago, before the analysis had taken the direction it's taken. Needs to be rewritten. Focus on trajectories of recovery. Focus on differences across population groups. Talk about remaining un-addressed biases in the data. With our methodology what we achieve is 1) data imputation for missing counts in regions that have low counts due to a variety of reasons -discuss- 2) data adjustment for fluctuations in daily number of users. However, our method does not explicitely adjust underrepresentation of specific population groups. This remains as future work and we should acknowledge the importance of this to make research based on digital traces more relevant for policy.

Using location data from Meta-Facebook users, our study aimed to examine the evolution of patterns of mobility across socioeconomic groups in functional urban areas from Argentina, Chile, Colombia and Mexico from April 2020 to March 2023, following the COVID-19 pandemic. We found a systematic drop in the number of population movements in April 2020, with the largest reductions observed in the most affluent administrative units within functional urban areas (FUAs) from Argentina, Chile and Colombia. While mobility rebounded closer to pre-pandemic levels approximately two years later, when COVID-19 restrictions eased, the number of movements remained below pre-pandemic in Chile. Furthermore, we found that at

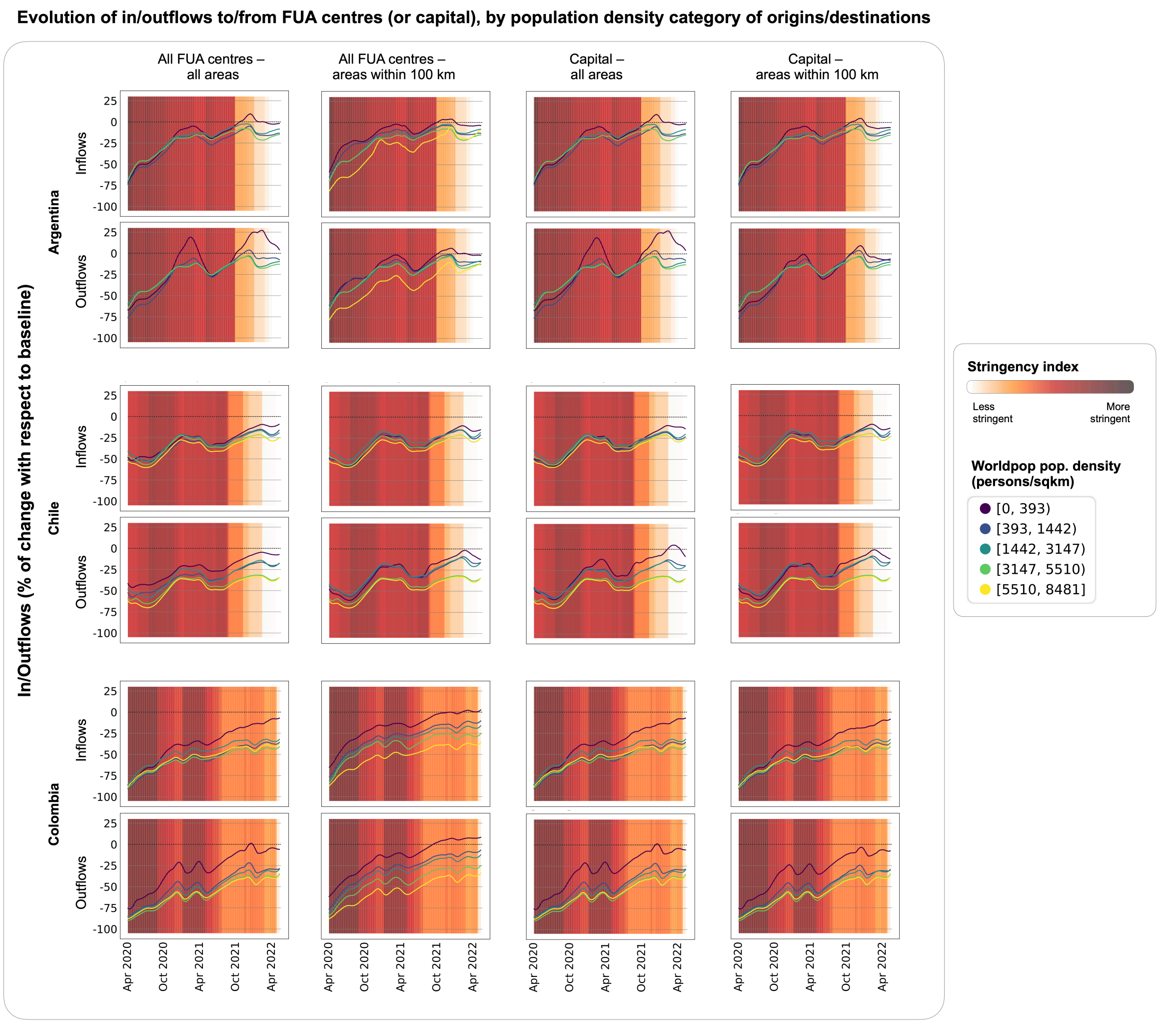

Figure 6: Evolution of number of inflows (and outflows) to (from) functional urban area centres, by population desnity category of origin (destiantion) for Argentina, Chile and Colombia, as percentage change with respect to pre-pandemic baseline levels.

the beginning of the pandemic there were inequalities between socioeconomic groups in terms of the levels of urban mobility. While it has taken more than two years for Argentina and Mexico to gradually reduce gap, inequalities persist as of March 2023, especially in Chile and Colombia according to our estimated data.

We focused the analysis on short-distance movements in urban areas, especifically those covering 70 km or less. These journeys are typically considered to represent local and routine mobility (Owen and Green 1992) . However, due to the characteristics of the Meta-Facebook movement data, we are unable to distinguish the purpose of these short-distance movements. Hence, some of our data could be capturing journeys that involve a permanent change of place of residence. Our work therefore motivates the need to answer questions regarding the validity of digital footprint data for the analysis of human mobility. Further research should focus on inferring more specific information about the nature of the journeys, following similar approaches to those proposed by Cabrera-Arnau et al. (2023), and quantifying the extent to which the digital footprint data mirrors the true mobility patterns.

Conducting research on urban mobility using digital footprint data is not straightforward, due to the challenges in accessing and working with unstructured data sets which are often subject to biases and statistical representation issues. These biases often arise from inequalities in access and usage of digital technologies across demographic groups (Rowe 2023a). Despite these challenges, the data and analysis that we used for this work provide evidence for non-trivial patterns that are consistent across four countries in Latin America and with other parts of the world. Our findings highlight the dynamic interplay between socioeconomic status and urban mobility, and shall be used to motivate and inform the public debate regarding the deep societal consequences of urban mobility disparities on the wider socioeconomic landscape of Latin American countries.

In conclusion, we argue that this work goes beyond the analysis of specific patterns by demonstrating the potential of digital footprint data for policy-relevant research on human mobility at an unprecedented level of spatiotemporal granularity. While we have seen a rise in initiatives to improve data services and methodological frameworks to facilitate the use of digital footprint data for social good, progress is still limited, especially in some parts of the world including Latin America. It is in the hands of governments and public organisations to prioritise the maximisation the societal benefits that digital footprint data has to offer. This includes engaging in activities such as building strategic partnerships with commercial data-holders and academic institutions to establish a unified framework for the use of digital footprint data in policy and research. In particular, we call for the creation of resources like those developed by the European Commission Joint Research Centre (Commission et al. 2022) and the UN Statistics Division (Division 2019), which identify sources of non-traditional data and set methodological protocols for incorporating mobile phone data into official mobility statistics. While current resources tend to have a global reach, we advocate for more tailored local initiatives that acknowledge disparities in regional data availability and adoption of digital technologies.

Data

Meta-Facebook data

Facebook Movements

To capture population movements, we used anonymised aggregate mobile phone location data from Meta users for Argentina, Colombia and Chile, covering a 24-month period from April 2020 to March 2022. We used the Facebook Movements datasets created by Meta and accessed through their Data for Good Initiative (https://dataforgood.facebook.com). The data are built from Facebook app users who have the location services setting turned on on their smartphone. Prior to releasing the datasets, Meta ensures privacy and anonymity by removing personal information and applying privacy-preserving techniques (Maas et al. 2019). One such technique involves adding a small, undisclosed amount of random noise to prevent the determination of precise counts for these areas. Another technique consists in applying spatial smoothing, using inverse distance-weighted averaging, to create a smoother data surface. Finally, small-count dropping is applied to exclude records where the data counts are below 10.

The Facebook Movements datasets provide information on the aggregated number of Facebook app users moving between pairs of locations, during the 24-month crisis period during COVID-19. The data is spatially aggregated into tiles according to the Bing Maps Tile System developed by Microsoft (rbrundritt). The Facebook Movement data for Argentina, Chile and Colombia are spatially aggregated into Bing tile levels 13… respectively *`@Carmen to check Bing Tile levels for movement data`*. That is about 4.9 x 4.9km at the Equator (Maas et al. 2019). The data is temporally aggregated into three daily 8-hour time windows which are 00:00-08:00, 08:00-16:00 and 16:00-00:00, Pacific Time (PT). These time windows translate to 05:00-13:00, 13:00-21:00, 21:00-05:00 for Argentina and Chile; and 03:00-11:00, 11:00-19:00, 19:00-03:00 for Colombia. For each time window, the origin location of a user is defined as the most frequently visited location in the previous time window, while the destination location is the most frequently visited location in the current time window.

In addition, each dataset includes baseline movement counts, which represent the estimated number of people moving prior to COVID-19. These baseline counts are calculated based on a 45-day period ending on March 10, 2020, using the average for the same time of day and day of the week within this period. For example, the baseline for the 00:00–08:00 (PT) window on Mondays is calculated by averaging all Monday records for that time window across the 45-day period. Further details about the baseline can be found in (Maas et al. 2019).

Due to the small-count dropping privacy measure, movement counts during both the crisis and baseline periods are excluded if they fall below 10. However, even when counts are not reported, the datasets consistently provide the percentage change in movement counts relative to the baseline. This percentage change, denoted as $y^{\%}_{ijdt}$, is computed as follows (Maas et al. 2019):

$$y^{\%}_{ijdt} = \frac{y^c_{ijdt} - y^b_{ijdt}}{y^b_{ijdt} - \varepsilon} \times 100.$$

Here, y^c_{ijdt} and y^b_{ijdt} represent the crisis and baseline movement counts, respectively, between origin tile i to destination tile j on day d and during time window t. For the baseline values y^b_{ijwt}, the subindex w denotes the day of the week corresponding to day d, as the baseline values are recorded by weekday, not by specific date. A small constant ε, usually set to 1, is added to the denominator to avoid division by zero (Maas et al. 2019).

Facebook Populations

As part of our proposed data processing methodology, we also utilize the Facebook Population datasets. Like the Facebook Movements data, these datasets cover a 24-month period from April 2020 to March 2022 and provide information on the number of active Facebook users at specific locations during the crisis period.

The Facebook Population data is aggregated spatially into Bing tiles and temporally into 8-hour time windows. A user's location is defined as the most frequently visited location within each 8-hour time window. Similar to the Facebook Movements datasets, the Population datasets include baseline counts, calculated by averaging values for the same day of the week and time of day over the 45-day pre-pandemic baseline period.

The same privacy-preserving techniques are applied to the Facebook Population datasets. As a result, population counts during both the crisis and baseline periods are excluded if they fall below 10. However, the datasets consistently report the percentage change in active user counts relative to the baseline. For tile i, day d and time window t, the relative change in active user counts $p^{\%}_{idt}$ is computed as:

$$p^{\%}_{idt} = \frac{p^c_{idt} - p^b_{iwt}}{p^b_{iwt} - \varepsilon} \times 100,$$

where p^c_{idt} and p^b_{iwt} represent the crisis and baseline active user counts at tile i on day d and during time window t. Once again, the subindex w in baseline values p^b_{iwt} denotes the day of the week corresponding to day d, and ε is a small constant usually set to 1, added to the denominator to avoid division by zero (Maas et al. 2019).

In the remainder of the paper, we sometimes use expressions such as the "Facebook population" or the "population of Facebook users" to refer to the number of active Facebook app users.

Meta-Facebook data pre-processing

For our analysis, we filter the data to include only one time window per day: 08:00–16:00 Pacific Time (PT), which corresponds to 13:00–21:00 in Argentina and Chile, and 11:00–19:00 in Colombia. This window is likely to capture the bulk of socioeconomic daytime activity, as well as a significant portion of morning commutes. The choice of considering only one time window per day is driven by the objective of analysing the evolution of movement patterns throughout the pandemic period, without focusing on variations through the day. By limiting the analysis to a single time window each day, we reduce potential noise and variability that could arise from considering multiple time windows per day. Furthermore, including all time windows could result in opposing movement trends that may cancel each other out when the data is aggregated by day. As a result of filtering the data to include only one time window per day, we can drop the subindex t from the variables defined above $y_{ijd}^{c,b,\%}$ and $p_{ijd}^{c,b,\%}$, as the time window is no longer a distinguishing factor in the analysis.

As a pre-processing step, we ensure that both the Facebook movement and population data are aligned in terms of spatial resolution for each country. Since the raw Facebook Population datasets are aggregated at a finer spatial resolution, we re-aggregate them to match the spatial resolution of the Facebook Movements data.

WorldPpop population data

We used data from WorldPop (Tatem 2017) to classify the spatial units of analysis according to their population density, and to estimate missing baseline values in the Facebook population data. WorldPop offers open access gridded population estimates at a resolution as small as 3 arc-seconds approximately 100m and 1km at the Equator, respectively. WorldPop produces these gridded datasets using top-down (i.e. dissagregating administrative area counts into smaller grid cells) or bottom-up (i.e. interpolating data from counts from sample locations into grid cells) approaches. For the purposes of this work, we use gridded population data at a resolution of 1km^2 in raster format. We perform a spatial join of the Facebook spatial units (Bing tiles) with the gridded population data and compute the sum of Worldpop populations corresponding to each of the Facebook spatial units.

Socioeconomic deprivation data

We use the Global Gridded Relative Deprivation Index (GRDI), Version 1 (GRDIv1) (Columbia University 2022) data set as a measure of socioeconomic deprivation. The GRDI data is made available via NASA's Socioeconomic Data and Applications Centre (SEDAC), at a spatial resolution of 30 arc-seconds, or 1 km^2 approximately. The index quantifies the relative levels of multidimensional deprivation and poverty, where a value of 100 represents the highest level of deprivation and a value of 0 the lowest. We perform a spatial join of the

Facebook spatial units and the gridded relative deprivation data and compute the average RDI corresponding to each of the Facebook spatial units.

Methods

`@Francisco to revise the whole section`

Processing Facebook data

A significant challenge in analysing population counts and movements using Facebook user data is the absence of records for small counts. This limitation stems from privacy-protection measures designed to prevent the identification of individuals or small groups based on their location. The missing data in the Facebook Movements and Facebook Population datasets are not distributed randomly. Spatially, these missing values display high spatial autocorrelation as shown in Supplementary Figure **?@fig-autocorrelation** for the Facebook Population datasets. Furthermore, due to the non-random nature of these missing values, spatial units that are sparsely populated or that have a low number of Facebook app users could potentially be underrepresented in the analysis. Therefore, simply removing the missing records from the analysis could lead to geographically biased results (Afghari et al. 2019). To address this, we designed a data processing method for missing data imputation. An overview of this data imputation method is provided in Figure 7.

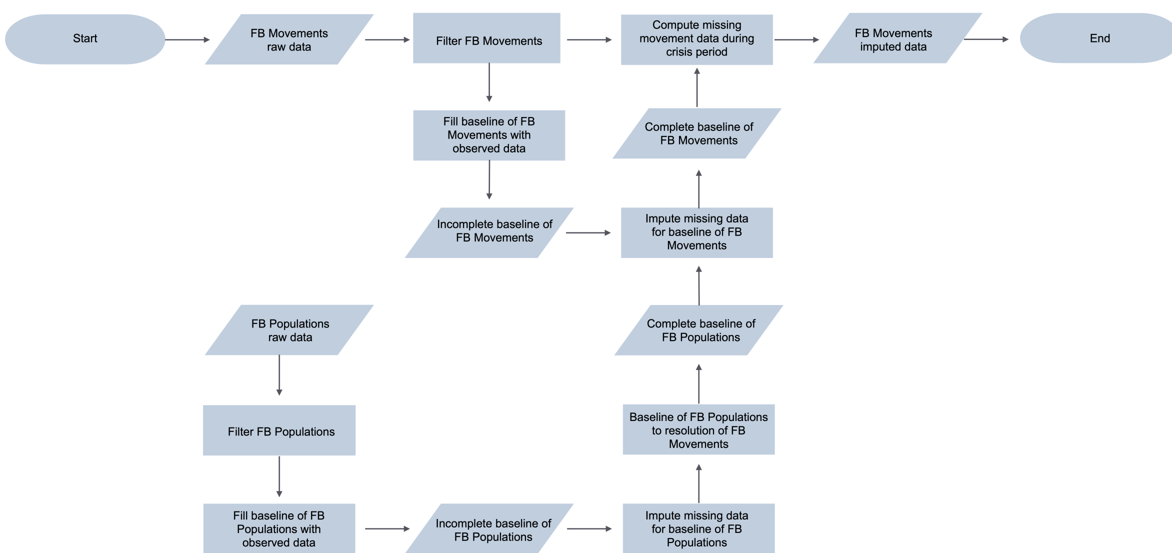

Figure 7: Workflow diagram for data imputation method.

Additionally, we apply correction factors and time series smoothing to eliminate the fluctuations in the daily number of observations, assuming that the representativeness of Facebook data across spatial units remains stable during the study period. The steps of our data-processing workflow are further described in the next four subsections.

Imputing Facebook Population data

We begin by identifying all the baseline values reported in the Facebook Population datasets, which form the basis for creating a new dataset that includes all available baseline information. Next, following (Duan et al. 2024), we use linear models to estimate the missing Facebook population baseline counts based on Worldpop population data. These models are fitted using ordinary least squares regression, with Worldpop data as the explanatory variable and the available baseline counts from the Facebook Population dataset as the dependent variable. The estimation process and its accuracy are illustrated in Supplementary Figure **?@fig-imputbaselinepop**.

We then use the complete baseline of Facebook population counts to compute missing Facebook population counts during the crisis period. This is possible because, as mentioned above, Meta reports the percentage change in the number of counts with respect to the baseline $p_{id}^{\%}$, even if the counts are not reported due to low value.

Imputing Facebook movement data

The imputation of Facebook movement baseline values is done according to a spatial interaction model (see e.g. (Rowe, Lovelace, and Dennett 2022)). We consider baseline movement counts y_{ijw}^{b} between pairs of origin i and a destination j tiles on weekday w, and model this variable as a function of the Facebook population count at the origin tile on the same weekday, the Facebook population count at the destination on the same weekday and the distance between origin and destination. We also include indicator variables to capture the day of the week. Mathematically, this model can be expressed as

$$\mu_{ijw}^{b} = \beta_0 + \beta_1 p_{iw}^{b} + \beta_2 p_{jw}^{b} + \beta_3 d_{ij} + \beta_4 w + \varepsilon \tag{1}$$

where $\mu_{ijw}^{b} = E[y_{ijw}^{b}]$ is the expectation of the flow of people from tile i to tile j on the weekday w during the baseline period; β_0 is an intercept, p_{iw}^{b} and p_{jw}^{b} are the Facebook population counts at the origin and destination on weekday w during the baseline period, d_{ij} is the distance between origin and destination, w is a series of indicator variables capturing the day of the week, and $\beta_{0,1,2,3,4}$ are model parameters to be estimated from the observed data. The error term is denoted by . To estimate the model parameters, we used a Gaussian regression model, taking the log of the population at origin and at destination, and the log of the distance. Residual plots for the spatial interaction model are provided in the Supplementary Figure **?@fig-residualsim**.

We compute missing Facebook movement counts during the crisis period by considering the complete Facebook movements baseline and the percentage change in the number of counts with respect to the baseline, which is reported in the Facebook Movements datasets even when the count is not reported due to its low value.

Applying correction factors

The total number of active users within a given time window exhibits daily fluctuations due to limitations in Internet connectivity and user data access options (Maas et al. 2019). Following (Yabe et al. 2020) and (Duan et al. 2024), we applied a correction factor to mitigate the potential impact of these daily fluctuations on the results, since they could mask the mobility trends. This approach assumes that the representativeness of Facebook data across spatial units remains consistent throughout the study period.

The adjusted number of movements between tile i and tile j on day d, y'_{ijd} was obtained as:

$$y'_{ijd} = k_d \times y^c_{ijd}$$

where y^c_{ijd} is the original number of movements between tiles i and j on day d during the crises period and k_d is a correction factor computed as the median of the sum of active user counts across all days d divided by the sum of active user counts across all spatial units on day d. Mathematically,

$$k_d = \frac{\mathrm{med}_d\left(\sum_i p^c_{id}\right)}{\sum_i p^c_{id}}.$$

Time series smoothing

A time series for each pair of origin-destination tiles was generated using data processed as described above. However, the resulting time series still contained missing values due to days when no data was reported for specific location pairs. To ensure continuity, we addressed this issue by imputing missing values within the time series, replacing them with the average of the nearest 15 observations within the time series. This number was chosen to provide an optimal balance between maintaining temporal proximity and ensuring sufficient data coverage.

Additionally, to reduce noise and highlight underlying trends that might be obscured by short-term fluctuations, we applied a rolling-average smoothing technique to the resulting time series. This approach was used in the time series presented in Figure 2, Figure 3 and Figure 6.

Classification of tiles according to level of urbanisation and socioeconomic deprivation

The geographic distribution of categories for each of the countries is displayed in Figure 8 for Argentina, Chile and Colombia respectively. Figure 8 also shows, for each country, the proportion of the population across the various analysis categories is presented, based on both Facebook population counts and Worldpop population estimates. The Facebook population counts reflect the average number of active users across all weekdays in the pre-pandemic baseline period.

In all three countries, notable discrepancies appear between the population distributions according to WorldPop and Facebook data. For instance, in Argentina and Chile, WorldPop data indicates a higher proportion of people living in low-density areas, suggesting that Facebook data underrepresents populations in these regions. Similarly, in these countries, Facebook data shows an overrepresentation of the most affluent socioeconomic group. While addressing this kind of representativity bias is beyond the scope of this paper, we recognise its significance and the importance of addressing it in future analyses. This issue is revisited in the Discussion section.

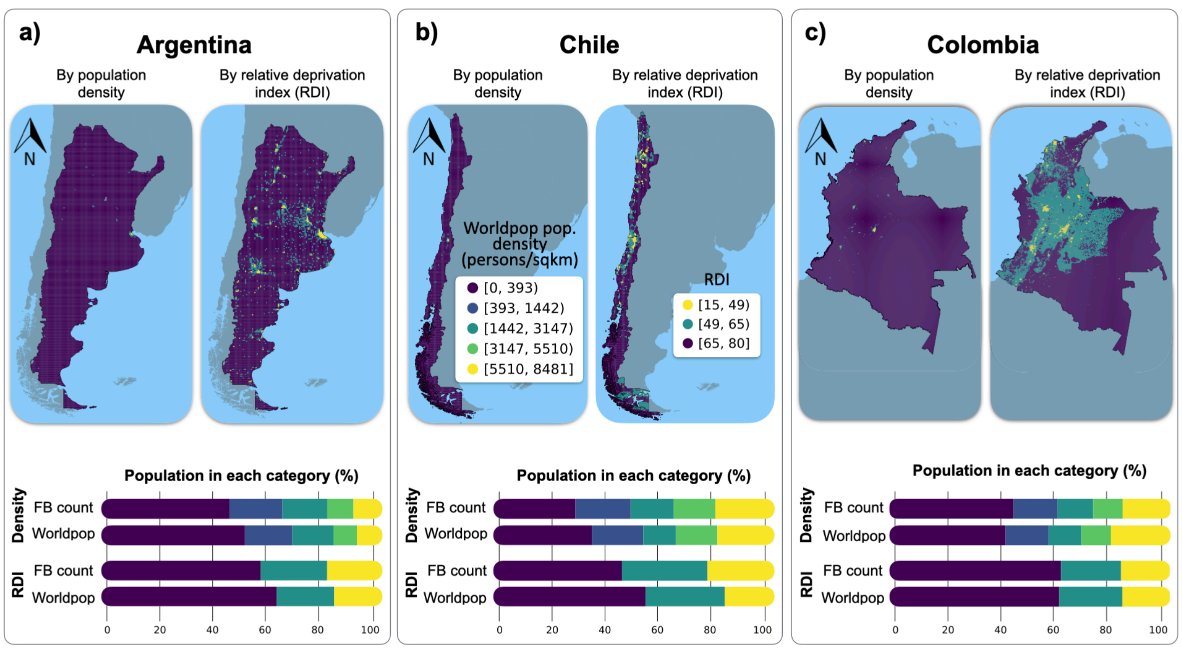

Figure 8: Classification of spatial units into categories by population density and by relative deprivation index. Spatial distribution of categories and population share in each category, by country.

Trend analysis

In order to quantify intergroup differences in the evolution towards pre-pandemic mobility patterns, we model the time series displayed in Figure 2 and Figure 3. To this end, we extract the trend, seasonal, and noise components of each time series using the `seasonal_decompose()` method from the time-series models and methods API in Python's `statsmodels` package (version 0.14.4). We then model the trend component according to five linear mixed-effects model specifications, using R's `glmmTMB` library (version 1.1.10).

Model 1 includes time as the sole explanatory variable, while Model 2 incorporates spatial heterogeneity by adding an indicator variable for the population density or RDI category of the origin. Models 3, 4, and 5 include time as a fixed effect but also incorporate random effects. Model 3 accounts for a random intercept based on the origin category, Model 4 includes a random slope for the origin category, and Model 5 combines both a random intercept and a random slope by origin category. Figures Figure 4 and Figure 5 illustrate the random variation in intercept and slope estimated by Models 3, 4, and 5. Full details on the parameter estimates are provided in Supplementary Tables (**tab-trendanalysis?**) for each country.

Code availability

For all studies using custom code in the generation or processing of datasets, a statement must be included under the heading "Code availability", indicating whether and how the code can be accessed, including any restrictions to access. This section should also include information on the versions of any software used, if relevant, and any specific variables or parameters used to generate, test, or process the current dataset.

References

Acknowledgements

The work was supported by Research England..., project no. ... , 'RECAST'.

Author information

Authors and affiliations

Contributions

A.A. conceived the experiment(s), A.A. and B.A. conducted the experiment(s), C.A. and D.A. analysed the results. All authors reviewed the manuscript.

Ethics declarations

Competing interests

The authors declare no competing interests.

Supplementary information

Rights and permissions

Abdullah, Muhammad, Charitha Dias, Deepti Muley, and Md. Shahin. 2020. "Exploring the Impacts of COVID-19 on Travel Behavior and Mode Preferences." *Transportation Research Interdisciplinary Perspectives* 8 (November): 100255. https://doi.org/10.1016/j.trip.2020.100255.

Abu-Rayash, Azzam, and Ibrahim Dincer. 2020. "Analysis of Mobility Trends During the COVID-19 Coronavirus Pandemic: Exploring the Impacts on Global Aviation and Travel in Selected Cities." *Energy Research & Social Science* 68 (October): 101693. https://doi.org/10.1016/j.erss.2020.101693.

Ackers, Louise. 2005. "Moving People and Knowledge: Scientific Mobility in the European Union1." *International Migration* 43 (5): 99–131. https://doi.org/10.1111/j.1468-2435.2005.00343.x.

Afghari, Amir Pooyan, Simon Washington, Carlo Prato, and Md Mazharul Haque. 2019. "Contrasting Case-Wise Deletion with Multiple Imputation and Latent Variable Approaches to Dealing with Missing Observations in Count Regression Models." *Analytic Methods in Accident Research* 24 (December): 100104. https://doi.org/10.1016/j.amar.2019.100104.

Aksoy, Cevat Giray, Jose Maria Barrero, Nicholas Bloom, Steven Davis, Mathias Dolls, and Pablo Zarate. 2022. "Working from Home Around the World." https://doi.org/10.3386/w30446.

Alexander, Lauren, Shan Jiang, Mikel Murga, and Marta C. González. 2015. "Origin–Destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data." *Transportation Research Part C: Emerging Technologies* 58: 240–50. https://doi.org/https://doi.org/10.1016/j.trc.2015.02.018.

Barbosa, Hugo, Marc Barthelemy, Gourab Ghoshal, Charlotte R. James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J. Ramasco, Filippo Simini, and Marcello Tomasini. 2018. "Human Mobility: Models and Applications." *Physics Reports* 734 (March): 1–74. https://doi.org/10.1016/j.physrep.2018.01.001.

Barrero, Jose Maria, Nicholas Bloom, and Steven Davis. 2021. "Why Working from Home Will Stick." https://doi.org/10.3386/w28731.

Belik, Vitaly, Theo Geisel, and Dirk Brockmann. 2011. "Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases." *Physical Review X* 1 (1). https://doi.org/10.1103/physrevx.1.011001.

Bell, Martin, Elin Charles-Edwards, Dorota Kupiszewska, Marek Kupiszewski, John Stillwell, and Yu Zhu. 2014. "Internal Migration Data Around the World: Assessing Contemporary Practice." *Population, Space and Place* 21 (1): 1–17. https://doi.org/10.1002/psp.1848.

Bonaccorsi, Giovanni, Francesco Pierri, Matteo Cinelli, Andrea Flori, Alessandro Galeazzi, Francesco Porcelli, Ana Lucia Schmidt, et al. 2020. "Economic and Social Consequences of Human Mobility Restrictions Under COVID-19." *Proceedings of the National Academy of Sciences* 117 (27): 15530–35. https://doi.org/10.1073/pnas.2007658117.

Borkowski, Przemysław, Magdalena Jażdżewska-Gutta, and Agnieszka Szmelter-Jarosz. 2021. "Lockdowned: Everyday Mobility Changes in Response to COVID-19." *Journal of Transport Geography* 90 (January): 102906. https://doi.org/10.1016/j.jtrangeo.2020.102906.

Cabrera-Arnau, Carmen, Chen Zhong, Michael Batty, Ricardo Silva, and Soong Moon Kang. 2023. "Inferring Urban Polycentricity from the Variability in Human Mobility Patterns." *Scientific Reports* 13 (1): 5751. https://doi.org/10.1038/s41598-023-33003-7.

Calafiore, Alessia, Krasen Samardzhiev, Francisco Rowe, Martin Fleischmann, and Daniel Arribas-Bel. 2023. "Inequalities in Experiencing Urban Functions. An Exploration of Human Digital (Geo-)footprints." *Environment and Planning B: Urban Analytics and City Science*, November. https://doi.org/10.1177/23998083231208507.

Carranza, Rafael, Mauricio De Rosa, and Ignacio Flores. 2023. "Wealth Inequality in Latin America."

CEU. JRC. 2019. *GHSL-OECD functional urban areas: public release of GHS FUA*. LU: Publications Office. https://doi.org/10.2760/67415.

Chang, Serina, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. 2020. "Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening." *Nature* 589 (7840): 82–87. https://doi.org/10.1038/s41586-020-2923-3.

Chen, Cynthia, Jingtao Ma, Yusak Susilo, Yu Liu, and Menglin Wang. 2016. "The Promises of Big Data and Small Data for Travel Behavior (Aka Human Mobility) Analysis." *Trans-*

portation Research Part C: Emerging Technologies 68 (July): 285–99. https://doi.org/10.1016/j.trc.2016.04.005.

Chinazzi, Matteo, Jessica T. Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, et al. 2020. "The Effect of Travel Restrictions on the Spread of the 2019 Novel Coronavirus (COVID-19) Outbreak." *Science* 368 (6489): 395–400. https://doi.org/10.1126/science.aba9757.

Columbia University, Center for International Earth Science Information Network. CIESIN. 2022. "Global Gridded Relative Deprivation Index (GRDI), Version 1." Palisades, New York: NASA Socioeconomic Data; Applications Center (SEDAC). https://doi.org/10.7927/3xxe-ap97.

Commission, European, Joint Research Centre, C Bosco, S Grubanov-Boskovic, S Iacus, U Minora, F Sermi, and S Spyratos. 2022. *Data Innovation in Demography, Migration and Human Mobility*. Publications Office of the European Union. https://doi.org/doi/10.2760/027157.

De Ferranti, David M. 2004. *Inequality in Latin America: Breaking with History?* World Bank publications.

Division, United Nations Statistical. 2019. "Handbook on the Use of Mobile Phone Data for Official Statistics."

Duan, Qianwen, Jessica Steele, Zhifeng Cheng, Eimear Cleary, Nick Ruktanonchai, Hal Voepel, Tim O'Riordan, et al. 2024. "Identifying Counter-Urbanisation Using Facebook's User Count Data." *Habitat International* 150 (August): 103113. https://doi.org/10.1016/j.habitatint.2024.103113.

Dueñas, Marco, Mercedes Campi, and Luis E. Olmos. 2021. "Changes in Mobility and Socioeconomic Conditions During the COVID-19 Outbreak." *Humanities and Social Sciences Communications* 8 (1). https://doi.org/10.1057/s41599-021-00775-0.

Florida, Richard, Andrés Rodríguez-Pose, and Michael Storper. 2021. "Critical Commentary: Cities in a Post-COVID World." *Urban Studies* 60 (8): 1509–31. https://doi.org/10.1177/00420980211018072.

Fraiberger, Samuel P., Pablo Astudillo, Lorenzo Candeago, Alex Chunet, Nicholas K. W. Jones, Maham Faisal Khan, Bruno Lepri, et al. 2020. "Uncovering Socioeconomic Gaps in Mobility Reduction During the COVID-19 Pandemic Using Location Data." https://doi.org/10.48550/ARXIV.2006.15195.

Green, Mark, Frances Darlington Pollock, and Francisco Rowe. 2021. "New Forms of Data and New Forms of Opportunities to Monitor and Tackle a Pandemic." In, 423–29. Springer International Publishing. https://doi.org/10.1007/978-3-030-70179-6_56.

Habitat, UN. 2022. "World Cities Report 2022: Envisaging the Future of Cities." *United Nations Human Settlements Programme: Nairobi, Kenya*, 41–44.

Klugman, Jeni. 2009. "Human Development Report 2009. Overcoming Barriers: Human Mobility and Development." *Overcoming Barriers: Human Mobility and Development (October 5, 2009). UNDP-HDRO Human Development Reports*.

Lee, Sujin, Eunjeong Ko, Kitae Jang, and Suji Kim. 2023. "Understanding Individual-Level Travel Behavior Changes Due to COVID-19: Trip Frequency, Trip Regularity, and Trip Distance." *Cities* 135 (April): 104223. https://doi.org/10.1016/j.cities.2023.104223.

Maas, P., S. Iyer, A. Gros, W. Park, L. McGorman, C. Nayak, and P. A. Dow. 2019. "Facebook Disaster Maps: Aggregate Insights for Crisis Response and Recovery." In *16th International Conference on Information Systems for Crisis Response and Management*, 836–47.

Meng, Chuishi, Yu Cui, Qing He, Lu Su, and Jing Gao. 2017. "Travel Purpose Inference with GPS Trajectories, POIs, and Geo-Tagged Social Media Data." In *2017 IEEE International Conference on Big Data (Big Data)*, 1319–24. https://doi.org/10.1109/BigData.2017.8258062.

Milanovic, Branko. 2016. *Global Inequality: A New Approach for the Age of Globalization*. Harvard University Press.

Nilufer Sari Aslam, Tao Cheng, Mohamed R. Ibrahim, and Yang Zhang. 2021. "ActivityNET: Neural Networks to Predict Public Transport Trip Purposes from Individual Smart Card Data and POIs." *Geo-Spatial Information Science* 24 (4): 711–21. https://doi.org/10.1080/10095020.2021.1985943.

Nouvellet, Pierre, Sangeeta Bhatia, Anne Cori, Kylie E. C. Ainslie, Marc Baguelin, Samir Bhatt, Adhiratha Boonyasiri, et al. 2021. "Reduction in Mobility and COVID-19 Transmission." *Nature Communications* 12 (1). https://doi.org/10.1038/s41467-021-21358-2.

Owen, D., and A. Green. 1992. "Migration Patterns and Trends." In *Migration Processes and Patterns: Research Progress and Prospects*, edited by T. Champion and T. Fielding. Belhaven Press.

rbrundritt. "Bing Maps Tile System - Bing Maps — Learn.microsoft.com." https://learn.microsoft.com/en-us/bingmaps/articles/bing-maps-tile-system.

Rowe, Francisco. 2023a. "9.: Big Data." In *Concise Encyclopedia of Human Geography*, 42–47. Cheltenham, UK: Edward Elgar Publishing. https://doi.org/10.4337/9781800883499.ch09.

———. 2023b. "Big Data." In *Concise Encyclopedia of Human Geography*, 42–47. Edward Elgar Publishing.

Rowe, Francisco, Carmen Cabrera-Arnau, Miguel González-Leonardo, Andrea Nasuto, and Ruth Neville. 2023. "Reduced Mobility? Urban Exodus? Medium-Term Impacts of the COVID-19 Pandemic on Internal Population Movements in Latin American Countries." https://doi.org/10.48550/ARXIV.2311.01464.

———. 2024. "Medium-Term Changes in the Patterns of Internal Population Movements in Latin American Countries: Effects of the COVID-19 Pandemic." In *Population and Development Series*, No. 139 (LC/TS.2024/71). Santiago: Economic Commission for Latin America; the Caribbean (ECLAC).

Rowe, Francisco, Alessia Calafiore, Daniel Arribas-Bel, Krasen Samardzhiev, and Martin Fleischmann. 2022. "Urban Exodus? Understanding Human Mobility in Britain During the COVID-19 Pandemic Using Facebook Data." https://doi.org/10.48550/ARXIV.2206.03272.

Rowe, Francisco, Miguel González-Leonardo, and Tony Champion. 2023. "Virtual Special Issue: Internal Migration in Times of COVID-19." *Population, Space and Place*, March. https://doi.org/10.1002/psp.2652.

Rowe, Francisco, Robin Lovelace, and Adam Dennett. 2022. "Spatial Interaction Modelling: A Manifesto." http://dx.doi.org/10.31219/osf.io/xcdms.

Santana, Clodomir, Federico Botta, Hugo Barbosa, Filippo Privitera, Ronaldo Menezes, and

Riccardo Di Clemente. 2023. “COVID-19 Is Linked to Changes in the Time–space Dimension of Human Mobility.” *Nature Human Behaviour*, July. https://doi.org/10.1038/s41562-023-01660-3.

Schiavina M., Maffenini L., Moreno-Monroy A. 2019. “GHS-FUA R2019A - GHS Functional Urban Areas, Derived from GHS-UCDB R2019A, (2015), R2019A.” European Commission, Joint Research Centre (JRC).

Tatem, Andrew J. 2017. “WorldPop, Open Data for Spatial Demography.” *Scientific Data* 4 (1). https://doi.org/10.1038/sdata.2017.4.

Taylor, Sean J, and Benjamin Letham. 2017. “Forecasting at Scale.” *PeerJ Preprints* 5 (September): e3190v2. https://doi.org/10.7287/peerj.preprints.3190v2.

United Nations, Department for Economic, and Social Affairs. 2023. *World Population Prospects 2022: Summary of Results*. UN.

Wang, Yikang, Chen Zhong, Qili Gao, and Carmen Cabrera-Arnau. 2022. “Understanding Internal Migration in the UK Before and During the COVID-19 Pandemic Using Twitter Data.” *Urban Informatics* 1 (1). https://doi.org/10.1007/s44212-022-00018-w.

Weill, Joakim A., Matthieu Stigler, Olivier Deschenes, and Michael R. Springborn. 2020. “Social Distancing Responses to COVID-19 Emergency Declarations Strongly Differentiated by Income.” *Proceedings of the National Academy of Sciences* 117 (33): 19658–60. https://doi.org/10.1073/pnas.2009412117.

Yabe, Takahiro, Kota Tsubouchi, Naoya Fujiwara, Takayuki Wada, Yoshihide Sekimoto, and Satish V. Ukkusuri. 2020. “Non-Compulsory Measures Sufficiently Reduced Human Mobility in Tokyo During the COVID-19 Epidemic.” *Scientific Reports* 10 (1). https://doi.org/10.1038/s41598-020-75033-5.