Building Geospatial Health Solutions in Emerging Markets using *Meta Data for Good* Anonymized Human Mobility Dataset

Olubayo Adekanmbi PhD
CEO/Lead Consultant
Data Science Nigeria/DSNai (Nigeria and United Kingdom)
olubayo@dsn.ai

Table of Contents

1. Acknowledgments

Data Science Nigeria and DSNai would like to acknowledge Meta (formerly Facebook) for making the Data for Good suite available for public use. The dataset provided the foundation for all the various use cases, theoretical explorations and geospatial experimentations applied in this white paper with hands-on delivery on specific projects across four African countries (Kenya, Nigeria, Senegal and South Africa). Our profound gratitude also goes to the Africa GeoPortal platform (managed by ESRI), for providing a free platform of geospatial tools, data, and training for users working on Africa's geospatial challenges. We appreciate WorldPop for its open source datasets, API and tools which enhanced our use cases development through the application of high resolution, contemporary spatial data on human population distributions, their characteristics and spatio-temporal changes over time to support health-related solution design. We appreciate the effort of our multi-country delivery partners across the four African countries including everyone who was involved in the content development, hackathon, workshops and hands-on use case exploration.

We appreciate effort of the DSN multi-campus learning communities who mobilized their students and community members to participate in the hackathon and learning workshops. We specially acknowledge the opportunities given to us by global development agencies and organizations like the Bill and Melinda Gates Foundation, Mastercard Foundation, World Bank, DataDotOrg, AI Commons, Microsoft, Google.org, Flowminders, University College London and TikTok (ByteDance Limited) to creatively explore how the power of precision data can drive sustainable socio-economic development at scale. This endeavor would not have been possible without the support of the entire DSN team who readily provided guidance, encouragement, tremendous amount of time, and necessary resources to ensure that this whitepaper sees the light of the day.

2.

Introduction to Mobility Data

1.1 Mobility Data: What is it?

Taking a long view of human history, some innovations rise and are eventually replaced by new ones, whereas other technologies take hold yet keep evolving over time. When it comes to the science of information and its place in data-driven decision-making pipelines, the latter is the case. In order to satisfy basic needs, sedentary humans had to decide where to grow their food, where to get water, where to live, etc. But soon they will realize that what was at the where of yesterday may not be found at the where of today. Hence, the start of the quest for time-based information (the when).

The centuries long dependence on this kind of information with one's surrounding, growing from a single village to the whole globe coupled with the invention of densely networked computer systems, eventually led to the advent of Geographical Information Systems (GIS). As a result, humans can better coordinate to manage their natural resources, and indeed the very first GIS enabled a land-use management plan at country level [1]. With the rise of computing power and satellite imaging, the ambitious use and expansion of earth-based location data, also known as geospatial data, increased. This aspiration becomes evident through the shift of interest in spatio-temporal information from static bodies such as streams and land masses, to moving bodies such as people and animals hence the birth of mobility data.

In the recent Dagstuhl Seminar "Mobility Data Science", participants from industry and academia agreed to define Mobility Data as "Spatio-temporal data capturing behavior of moving entities" [2]. With this target in hand, questions then arise about the added advantage of capturing moving entities behavior over that of static entities.

Beyond scientific advancement, what utility can be gained from the introduction of temporally as well as spatially resolved data? Mobility data science is not merely the tracking of the" where" and" when" of the moving entity. For such data to be useful, it is necessary to turn raw data into useful information leading to impactful knowledge and ultimately to socially contextualized wisdom. Only by effectively merging all these components of data science and embedding the scientific information into domains of social activity can humanity provide structure to and unlock the value latent within the vast array of data made available by modern information technology.

Our goal in this paper is to review one particular domain of data that has grown in quantity, importance, and usability in recent decades: mobility data. In particular, we will focus on translational applications of data science techniques to the *Data For Good* data suite, made publicly available by

Facebook/Meta, which brings together data sources from social media usage, satellite imaging, as well as geo- and demographic data into a diverse collection of datasets that can be used to generate insight into health, economics, migration, and a variety of other human phenomena.

1.2 Geospatial Terms and Acronyms

Spatial Terminologies

Terms	Definition
Spatial Data	Data associated with space/location. A more precise term is Geospatial Data, where the location is referenced to a point on the earth's surface.This data is obtained from the observation of events or measurements of an object using GPS (Global Positioning System) or mobile device, and it is expressed in the projected coordinate system (x, y). Other terms associated with this are, geospatial data, geometric data, geographic data, GIS data
Vector data	Data models are used to represent features of the earth with discrete geometric points, lines, and polygons, giving it a clear and defined boundary. Each feature is associated with attributes of various data types that describe the object.
Raster data	Another type of spatial data modelled to represent information as a continuous field over a space. In simple form, raster data is an image consisting of gridded pixels, where each pixel is a reference to a location on earth, and the numeric values are used to describe the earth's features. This is the most common output from satellites and remote sensors.

Coordinate	Geo-coordinate/coordinate references point/position of an object on the earth surface. It mostly consists of a pair of number represented as Latitude and Longitude (x, y) geometric point on the spherical earth space.
Geotag	This involves adding geographic information to a file content (text, photography, files, music, videos, e.t.c). This is a very common approach in data collection on the web and mobile devices.
Co-ordinate reference system	A coordinate reference system (CRS) defines how to convert 3-D spherical Map to a 2-D digital projected map while still referencing the real places on the earth. The type of CRS used in analysis and operation is highly dependent on the region of interest.
Zonal statistics	Zonal statistics applies statistical function based on defined region or zone within an area. This is often employed in summarizing raster data attribute over a given region, in most case the regions as defined in a polygon data.
Buffering	In geospatial processing, buffering means creating an outward zone of specific distance around a feature/object. This is a very efficient process used to understand proximity of an object with respect to other objects. For example, a hospital might want to know the number of accidents within 1.5 km distance of its location. Using buffering is one of the widely used function for neighborhood proximity.
Spatio-temporal	The representation of spatial data in space and time. In a spatio-temporal environment, data are dynamic over space and time. It is a crucial concept in understanding epidemics spread over time at different locations in the health sector.

Shapefile	Shapefile is the most common file format used to store vector geospatial data with point, lines, and polygon. Shapefile supports wide range of GIS software and can be used for analysis.
Spatial Relationship	Relationship among spatial data that defines how they are position in alignment to one another for better analysis. Examples are those which includes overlay, intersect, contains etc.
Web Mapping	The process of utilizing maps and spatial analysis for end-user consumption on the web via a browser. This is very common among the google map service, Bing and open street map.
Tessellation	Tessellation is the geometric process of partitioning a region into smaller units, mostly of same size and shape, that do not overlap but completely fill the entire area. The basic unit of tessellation can be squares, triangles, and hexagons.

Earth Observation terminologies

Terms	Definition
Crowdsourcing	This involves obtaining work, information, and data from a large group of people, either paid or unpaid, majorly over the internet. It is a good basis to get local and diverse information from different community.
Differential Privacy	This is a mathematical algorithm that aggregates data to extract behavioral pattern from a group while withholding information about unique individuals.
Bing Tiles	Square area of predefined shape that appear in Microsoft bing map services dividing the map of the world into square area grid. It is efficient for easy retrieval and display of small unit of map area.
Baseline Period	Baseline is a specific time frame that is used as a reference for comparison. This helps to understand changes and trends in specific location using the base period which are usually a single day as a reference point. Majority of the baseline period are time before the lockdown period of COVID.
Trend Graph	This is a type of graph used to show data over a period to discover and disclose pattern in the data. A typical trend graph shows time in the horizonal (x) axis and values representing data are shown in the vertical (y) axis with a line joining the different data point in time.

3.

Theoretical Concepts in Human Mobility

3.1 Movement Sequences & Movement Trajectories

When moving, humans go through indefinite number of points. For the sake of data analysis, only "meaningful" points are considered. These are also known as locations or stay-points. The consecutive recording of these specific points is called a "movement sequence". E.g.: A sequence S= [A, B, A] tells that a person left a location A to go to a location B and then came back to location A.

A movement trajectory is a subset of the movement sequence. In the sequence [A, B, A], there are trajectories [A, B] and [B, A] for example.

3.2 Mobility Signature

Human communication is not spread equally to all known people but allocates most of the time and interest to specific people such as family and friends, resulting in a social signature. Similarly, collective mobility from a city is not equally distributed to all regions within a country but some regions are preferred over others resulting hence in a mobility signature.

These mobility signatures are location-specific and sort the different locations from the most visited one to the least[3]. This is achieved by counting all trips from a specific origin to various destinations and normalize this count by the total number of trips.

How is it useful? Mobility signatures quantify mobility patterns relative to overall mobility network, giving the relative importance of different cities visited. Thus, the change in mobility signature can be used as a proxy for economic development, demographic growth, etc.

3.3 Activity Space Maps

One of the great advantages of digitally-collected mobility data over traditional movement data sets such as OD (Origin/Destination) matrix is the presence of the time dimension. In fact, there is much to learn from the amount of time one spends in a specific geographic location. This results in *Activity Space Maps*[4], a novel global-scale movement and mobility data set which describes how people distribute their time through geographic space. They give information about how people from same location move whether at daytime or night-time providing therefore information about their short trips or commuting for example.

How is it useful? Some diseases vectors thrive in specific environment such as the Anopheles mosquitoes which cause Malaria. Understanding the risk of being contaminated based on time people spent in high-risk areas is of utmost usefulness.

3.4 Human mobility predictability

For the purpose of traffic forecasting, urban planning, disease modelling, etc. there is a pressing need of predicting future locations of people. Different approaches to predict the next relevant location have been undertaken by researchers[5] such as:

- Next-time bin approach[6]: The locations are reported at regular time interval, and the prediction is for the next time interval. However, concerns arise as the prediction probability vary widely, especially due to self-transitions i.e. situations where the next location is similar to the previous one. In fact, this increases the predictability without necessary increasing the performance of the model.
- Next-place approach[7]: To respond to the bias induced by self-transitions, removing such locations and focusing only on the distinct locations in the sequence was proposed. This reduced the predictability value by about 25%.

As different researchers were working on different predictability models, they soon noticed that the predictability depended not only on the model used but also on the data set[5]. Therefore, they started looking for underlying factors within the data that affected the predictability. Here are presented a few that were discovered:

- Stationarity[8]: A sequence S1= [1, 1, 2, 2, 3, 3, 4, 4] is more predictable than [1,3,2,4,3,2,1,4] because it contains self-transitions.

- Regularity[8]: A sequence S1= [1,2,3,2,1,3,1,2] is more predictable than S2= [1,3,2,4,3,2,1,4] as there are unique 3 locations in S1 whereas there are 4 in S2.

- Diversity of trajectories[9]: A sequence S1= [1, 2, 3, 4, 1, 2, 3, 4] is more predictable than S2= [1,3,2,4,3,2,1,4] because there are repeating patterns also known as routines in S1.

3.5 Theoretical examples on the use of social media data

Human mobility does not only depend on human behavior but it also reveals it. Human mobility data unfold both the similarity [3] and variety [10] of mobility patterns.

In their research, Liao, Yeh, and Jeuken [10], using geotagged tweets, proposed the quantification of mobility patterns' heterogeneity based on two features: geographical characteristics (global vs local) and network properties (explorer vs returner).

The geographical characteristics focus on how far the individual travels whereas the network properties highlight how exploratory the individual's journey is. The results show that more than 90% travel locally (local explorers -78.0%, local returners -14.4%). On the other side, among those who adventure further, many prefer to decentralize their trips (global explorers -7.3%, global returners -0.3%).

How is it useful? This is greatly informative for tourism businesses, transport planning, urban planning, etc. Rightly, the researchers pointed out the limitation of generalizing these results to population given the population and behavior bias in the geotagged data. Identifying and correcting those biases is therefore a great scope for future research.

4.

Broader insights into the application of mobility data for health interventions

In the field of healthcare, human mobility data can provide valuable insights that can inform responses to public health emergencies, accurately model global as well as local spatial spread of infectious diseases, and even predict outbreak locations at early stages of contaminations.

4.1 Ensuring data availability during public health emergencies

In their white paper Human *Mobility Data in Public Health* Emergencies[11], Liao, Yeh, and Jeuken highlighted 3 main axes that determine the success of a data-informed public health intervention: Data readiness, Methods readiness and Translational readiness—three factors driving successful implementation. The challenge is to ensure that efforts from technology companies producing the data, researchers who analyse and interpret them, and response agencies who use the products , all converge to produce informed decisions and deliver practical line of actions.

Optimal **Data Readiness** means that the data should be available in a readily usable format in order to be processed for public health emergency responses. To achieve this, clear interoperability standards among technology companies, researchers and response agencies should be agreed upon in advance, rather than jury-rigged on the fly during an emergency scenario.

For example, reaching a consensus on the methods used in the aggregation and anonymization of data would represent an important milestone. Though technology companies are the collectors of much raw data, they do not in reality own them.

To mitigate ambiguity, the use and re-use of data should be governed by specific legislation and not solely by internal policies of the data holders.

Meanwhile, appointing data stewards representing all stakeholders who help to oversee the collection, sharing and use of data would already serve as a significant breakthrough. As producers of the mobility data and recipients of the public health responses, communities should be involved, right from the data collection stage, in the conceptual development of probable use cases, from which useful data formats flow.

In fact, the more the utility of mobility data in public health strategies is proven in actual studies, the more the production and delivery of these data is likely to increase.

Methods Readiness covers the state of the methods used to extract data to accurately respond to public emergencies. Knowing that technologies through which mobility data are collected are not equally distributed over a population, one of the challenges is data representativeness. This is of crucial importance for public health application as all people regardless of their access to recent technologies should equally have access to proper healthcare. For this reason, any statistical bias in the data should be identified, communicated, and taken into consideration in designing an appropriate public health response.

Of equal importance is the challenge of preserving privacy. Even though, individual information is mostly protected through the aggregation and anonymization process, even community-level information such as areas with the highest HIV infections can still undermine the personal privacy and security of members of the community. For this reason, wisdom should dictate a framework acknowledging the societal and political context of all data collected. Translational readiness refers to the integration of human mobility data into public health response pipelines and feedback loops. Indeed, the output of the data analysis ought to be integrated with other data sources such as population density, socio-economic development, etc, such as are available through governmental administrative units.

On one side, policymakers should guide researchers to prioritize producting actionable analytics. On the other side, networks of local scientists should support response agencies on the field. An even better scenario is the targeted training of a labor pool of “data bilinguals” i.e. people who understand both the data science and the policy implications in public health responses. Ultimately, the socialisation of the end products provides evidence of the utility of these data.

4.2 Tracking global spread of diseases

Thanks to the innovative means of transportation, humans have never travelled faster nor further in history. Now, with supersonic speed, an individual can reach the other side of the globe.

Unfortunately, this also means a more rapid and long distance spread of infectious diseases. Indeed, the higher the level of human mobility, the higher the mobility of any pathogen he may carry. For that reason, global health security, early warning systems and predictive models become imperative fast outbreak location.

The world has already suffered from the consequences of the unprecedented fast pathogens mobility like Influenza H1N1 (2009), MERS-CoV (2012), COVID-19 (2019), etc.[12].

IMAGE SOURCE:
GETTY IMAGE

4.3 Urban spatial patterns of infection

Traffic data have been used for a long time to model infection spread at national or international level. Mobility within cities being more complex and more difficult to capture, and so fewer attempts have been made to understand infection spatial distribution within urban areas.

However, thanks to the advent of high-resolution human mobility data such as from Call Detail Records (CDR) and GPS data, researchers such as Moss et al. [32] have worked on improving predictions of spatial patterns of influenza infection in Melbourne (Australia).

In their study, they assumed three (3) different spatial patterns for an individual to be infected:

- an individual's infection may occur in that individual's home location
- an individual may be contaminated by people living in visited locations
- an individual may be contaminated by people encountered within the hubs-and-spokes of public transport networks.

Their findings show that public transportation facilitated an "onion" spatial pattern of infection around transit network nodes, whereas private means of transport spread the infection further along in a North-West to South-East orientation. These findings illustrate the utility of incorporating mobility maps into disease spread modeling, prediction, and policymaking loops.

4.4 Finding disease outbreak locations ex-post

The determination of origin of disease outbreaks is vital to controlling/stopping its broader spread. Current methods such as contact tracing, transmission chain tracking, etc. are unfortunately not applicable at the early onset of an outbreak as no identified transmission chain or genomic data is available or identified at that early stage.

Schlosser and Brockmann [14] proposed a method to identify outbreak location(s) as the shared locations of affected individuals within a specific time-frame.

Although this requires high spatial resolution and non-anonymized data, they suggested a triangulation approach: requesting data directly from the identified affected individuals for the sake of public health benefits.

5.

Meta *Data for Good*

— Exploration and Explanation of Methodology

5.1 *Data for Good:* What is it?

Meta *Data for Good* (previously known as Facebook *Data for Good*) [15] started in 2017 with the objective to help Non-Governmental Organizations in their responses to natural disasters. Now, partners have been extended from humanitarian actors to academia, industries and even government to further support global development through research and informed policies[16].

Data and tools offered by Meta *Data for Good* are in three (3) different features as shown in Figure 2.1. In fact, Meta has access to data generated by around 3 billion of people.

Their data sources range from information from public posts, global survey's respondents from their Facebook platform to satellite imageries from external data sets. Whereas some of their products are publicly available, others are only shared on the basis of license agreements.

Although the access to both groups of datasets is free, privacy-preserving, and complies with data protection rules all over the world, the differences that characterize them are shown in Table 2.1[17].

WHAT WE OFFER

Data for Good builds maps, surveys, and insights.

Maps

From aggregated location insights to information on global friendship ties, we can build some pretty cool maps.

Surveys

The reach of the Facebook platform allows us to recruit participants for global surveys on a range of topics

Insights

Billions of public posts are shared on Facebook every months. These posts can reveal how communities feel about issues such as gender equality, climate and public healths

Figure 1.1: *Data For Good*: Tools & Data [15]

Public	Controlled
No license agreement required	License agreement required
Access through UN's Humanitarian Data Exchange Platform 1	Access through Meta's Partner Portal
Open to the Public	Only NGOs, academics and researchers

Table 1.1: *Data for Good*: Differences between Public and Private data

5.2 Population density & Population Flow

Facebook mobility data allocate users to specific tiles defined by the Bing Map Tiles system. Here are the primary data generated from the users' locations and movements based on an 8-hour window[18]:

- **population density**: Relative change of users within a tile
- **population flow**: Relative movement of users from one tile to another.

To preserve individual privacy and allow comparison with other datasets from different sources and information such as socio-economic or demographic data, these data are spatially aggregated from tile levels to administrative zones and temporally aggregated from 8-hour to 24-hour.

For example, in the Facebook movement range maps:

- all day ratio single tile users reflect the attribute of population density as it shows the positive proportion of users staying put within a single location;
- all day bing tiles visited relative change reflects the attribute of population flow as it shows the positive or negative change in movement relative to a baseline.

5.3 Understanding Meta *Data for Good*

Meta *Data for Good* provides seventeen types of datasets under six broad categories. Comprehensive guidelines such as methodology details, data features and tutorial packages can be consulted on the *Data for Good* website.

Below, we provide a short description of each DFG data tranche, which are discussed in detail in Section 3.

1. **Population**

- ***High Resolution Population Density Maps:*** This dataset is built upon satellite imageries and census data and is available for more than 160 countries.
Metrics: Population estimate at 30-meter resolution; Estimate of women, men, youth, children, women of reproductive age, and the elderly at the same resolution.
Users: Anyone with an interest population demographics.
Access type: Open to public

- ***Facebook Population During Crisis***
Metrics: Percent change in population in and around the disaster-affected area; Z-score expressing statistical significance of that change.
Users: Researchers or disaster response

agencies *(e.g. How has population increased in an area B consequently to a crisis in an area A?)*
Access type: Open to non-profits and researchers through a license agreement.

2. **Mobility**

- *Movement Range Maps:*
This dataset is available for more than 140 countries, and ready-made visualisations available for some of them.
Metrics: Change in movement of people, Number of people who stayed put in a single tile
Users: Researchers and Public health experts *(e.g. How do people respond to physical distancing measures?)*
Access type: Open to public

- *Facebook Movement During Crisis:*
Data set also available in interactive visualisation format
Metrics: Number of people moving from place to place within the delimited affected area in eight-hour increment
Users: Disaster response agencies
Access type: Open to non-profits and researchers through a license agreement.

- *Travel Patterns*
Metrics: Count of people moving from one country to another daily.
Users: Epidemiologists *(e.g. How international trips affect disease spread?)*, Economists *(e.g. How is economic development affected by travel patterns?)*, etc.
Access type: Open to non-profits and researchers through a license agreement.

- *Colocation Maps*
Metrics: Probability that two users from two different locations meet in one location
Users: Epidemiologists *(e.g. How disease spread from one region to another through human contact?)*
Access type: Open to non-profits and researchers through a license agreement.

- *Displacement Maps*
Metrics: Count, demographics and location *(city-level)* of people displaced as a consequence of a long-term crisis *(day fifteen post-crisis on-wards).*
Users: Researchers or Disaster response agencies *(e.g. When are people able to return home after a crisis?)*
Access type: Open to non-profits and researchers through a license agreement.

3. **Connectivity/Infrastructure**

- *Electrical Distribution Grid Maps:*
This data set is built upon a predictive model using satellite imageries. Combined with electrification investment scenarios, pathways to universal access can be assessed.
Metrics: Medium voltage electrical distribution infrastructure in any country.
Users: Policy makers and investors for infrastructure and community development projects, particularly in developing countries.
Access type: Open to public. However, model outputs are only available for six countries *(Malawi, Nigeria, Uganda, Democratic Republic of Congo, Côte d'Ivoire and Zambia).* Nevertheless, the documentation, code and even tutorial is available to replicate the model for your country of interest.

- *Network Coverage Maps*
  Metrics: Probability of network outages
  Users: Disaster response agencies ***(e.g. Where can we still communicate after a natural disaster?)***
  Access type: Open to non-profits and researchers through a license agreement.

- *Inclusive Internet Index*
  Metrics: Country overall score for internet use.
  Users: Policy makers and investors for internet connectivity projects.
  Access type: Open to public.

- *Infrastructure Route Study*
  Metrics: Cost effective routes for fiber network expansion
  Users: Policy makers and investors for internet connectivity, infrastructure and community development projects.
  Access type: Open to public. Available for Democratic Republic of Congo.

4. **Economic/Poverty**

- ***Relative Wealth Index***
  This dataset is built upon satellite imageries and connectivity data and is validated by ground truth measurements.
  Metrics: Standard of living of a micro-region ***(about 2.4 km2)*** relative to the rest of the country.
  Users: Humanitarian aid agencies, policy makers and other stakeholders for national development projects. Access type: Open to public

- ***Business Activity Trends:***
  This dataset is built upon business posting rates on Facebook.
  Metrics: Crisis -induced change in business activities per economic sector.
  Users: Researchers ***(e.g. When will an economic sector recover from a crisis?)***
  Access type: Open to non-profits and researchers through a license agreement.

- ***Commuting Zones***
  Metrics: Estimate of commuting zones at international level
  Users: Urban planners, Transportation engineers, etc.
  Access type: Open to public

5. **Social Connections**

- ***Social Connectedness Index***
  Metrics: Average scale of friendship ties between two geographical locations.
  Users: Economists ***(e.g. How are social connections linked to remittance flows in developing countries?)***
  Access type: Open to public

- ***Social Capital Atlas***
  Metrics: Network connectedness at school/ community level
  Users: Social, economic and even educational policy makers. ***(e.g. How social connections within educational institutions influence employment market?)***
  Access type: Open to public

6. **Forecasts**

COVID 19 Forecasts
Metrics: Prediction of COVID-19 spread in a specific area
Users: Healthcare management ***(e.g. Are the hospital's resources enough to cope with the incoming infection surge?)***, Public health experts ***(e.g. When shall we expect the next outbreak?)***
Access type: Open to public

- *Facebook Population During Crisis*
  Metrics: Percent change in population in and around the disaster-affected area; Z-score expressing statistical significance of that change.
  Users: Researchers or disaster response agencies ***(e.g. How has population increased in an area B consequently to a crisis in an area A?)***
  Access type: Open to non-profits and researchers through a license agreement.

5.4 Using *Data For Good*'s aggregated mobility data: the technical/analytical toolkit

In this section, we review the kinds of insights that can be extracted from aggregated mobility data, especially within the context of our work with Meta *Data For Good*.

Mobility data provide a treasure trove of aggregate statistics (e.g. geographically localized population densities, calling/texting volumes throughout the day/month/year) that can readily be extracted from call data in order to produce population-level analyses. Other applications require a substantial amount of data processing and augmentation in order to extract target insights.

For instance, an analysis of urban commuter mobility patterns at scale requires zooming in to establish fine-grained individual-level variables of interest—e.g. mapping individual device activity throughout the day into time/location-series transformed into imputed device trajectories—which can then be aggregated into measures of overall population flow.

This combination of both micro- and macro-level data routines is essential to making successful use of mobility data.

5.4.1 Aggregate-level analysis

Population data is an important aspect of understanding trends, and now with the availability of mobile device and the internet, it is easier to extract people's location and pattern based on regional activity from their mobile device.

The vast coverage and wide adoption of Facebook in Africa has rendered Facebook into a powerful tool to understand behavior changes in geographical region, which can be deployed in a variety of applications that are beneficial in addressing social issues.

In anonymized and aggregated form, the very same data collected from application usage and used to personalize and improve the user experience of a mobile application—e.g. to show relevant content and ads, recommend new social connections, etc.—can be utilized for population-level analyses in service of socially beneficial applications.

Mobility data and theirs uses can be categorized into five (5) different categories of spatial data structure and geographic resolution:

1. Population
2. Mobility
3. Connectivity/Infrastructure
4. Economic/Poverty
5. Social Connections

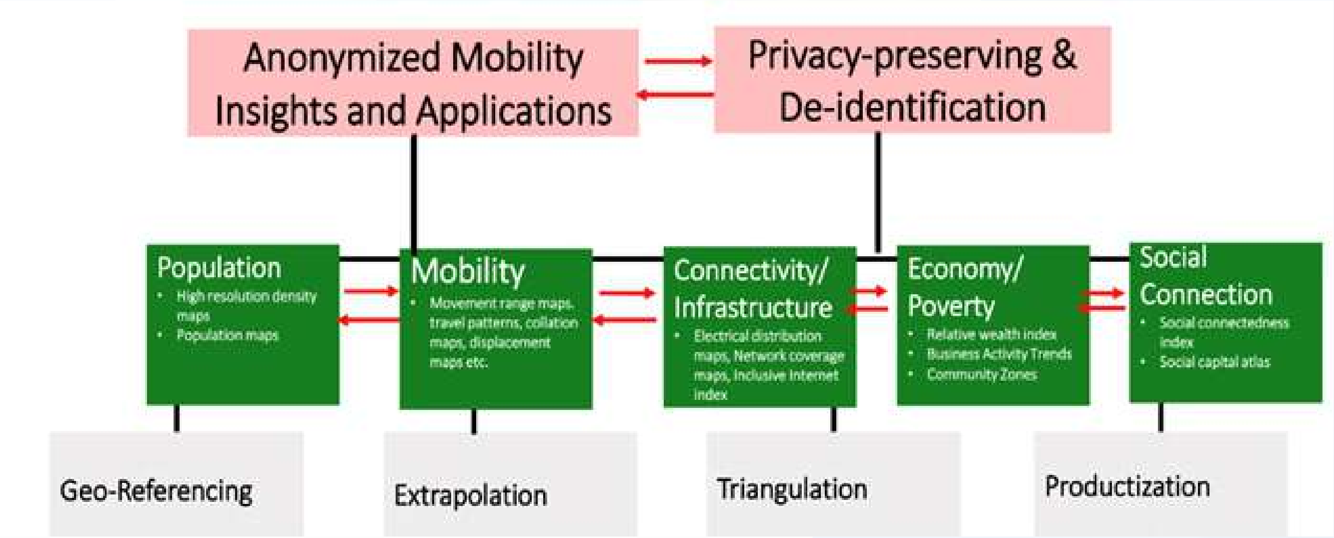


Figure 1.2 Meta *DataforGood* process

5.4.2 Population-level analysis

Population-level information are essential to many private applications and are also key in delivering social services, e.g. in the public and non-profit sector. There are two categories of population data technology available from the *Data For Good* repository.

First is the High-Resolution Population Density (HRPD), a map of population density data developed using computer-vision deep learning model to estimate building footprints from remote-sensed satellite imagery and census data. These density maps, available at 30m x 30m resolution for each covered population, also provide detailed imputations of different demographic weightings (women, men, youth children, women of reproductive age and elderly).

These data have been utilized in solving social issues such as infrastructure planning for vaccine campaign, disaster response efforts and risk analysis such as high accuracy flood risk analysis, and epidemic response in cases of COVID-19 and Dengue fever.

While the HRPD dataset do not draw on the rich resource of Facebook user data, the Facebook Population dataset is highly driven by user engagement in the Facebook software application. This data, efficient for disaster mapping, displays, among the Facebook user who have enabled Location Services, the user's location following a crisis compared to a pre-crisis baseline period at a regular interval of two weeks.

Individual locations in this dataset are mainly sampled within the spatial and temporal proximity of disaster events and highlight several key factors that help in understanding how populations are impacted by, preparing for, and affected by natural disasters.

In addition, this data acts as a good proxy for understanding population movement, eletricity/power availability, interruptions to network coverage, and migration flows caused by displacement. It has been utilized to understand the impact of natural hazard such as flood, bushfires, and cyclone Idai.

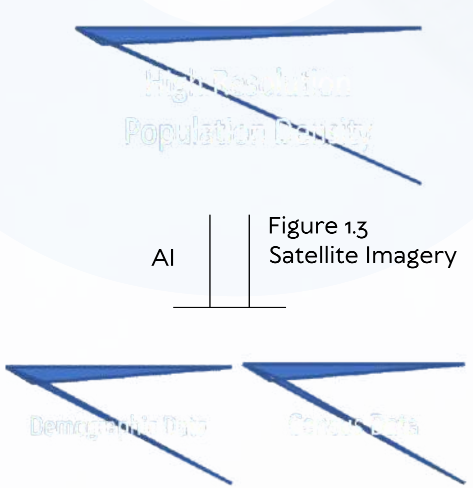


Figure 1.3

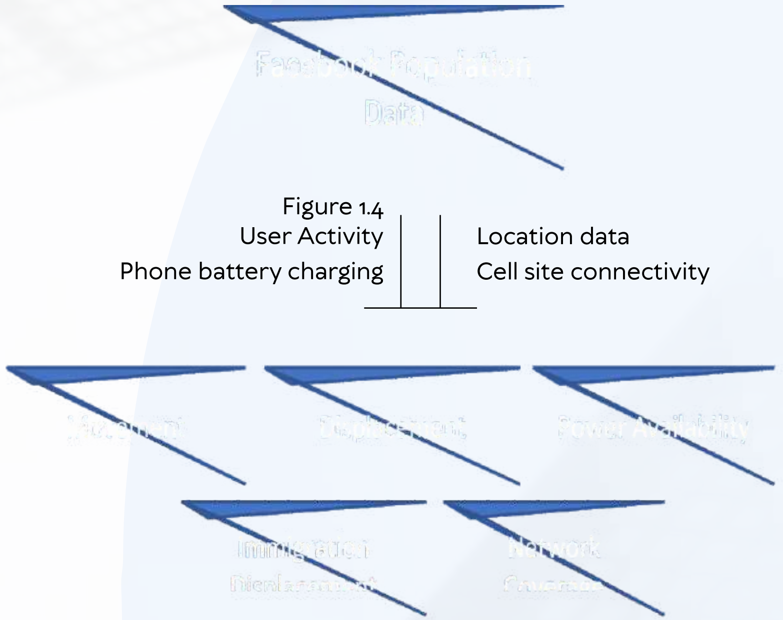


Figure 1.4

Resources:
https://dataforgood.facebook.com/dfg/tools/facebook-population-maps

https://dataforgood.facebook.com/dfg/tools/high-resolution-population-density-maps

Facebook Disaster Map
High Resolution Data

5.4.3 Analyzing mobility patterns

Temporally and spatially fixed network access samples grouped by individual, ordered by time, and linked into trajectories can provide high-resolution maps of individual and population flows. The Meta *Data For Good* mobility dataset represents an improvement to the static Facebook Population data by fusing it together with the High-Resolution Population Density dataset, allowing models describe the movement of people and generate insight from this collaborative development.

Within the Mobility segment, the most popular data are the movement range estimates containing metrics of change in movement rates, which can be used to understand movement trends and the tendency of populations staying in small areas during the entire day, using a baseline period for comparisons. This data is available daily for a given administrative boundary to a less aggregated level of Admin Boundary 3 divisions. However, regions with less activity and therefore sparse observation levels (using a threshold of 300 people) were exempted from the data, meaning that data are not available for some admin boundaries with less user Facebook activities. This can be important for respecting the privacy of users in rural residential zones who could be uniquely identified using these observations, but it is important to highlight that this restricts the insights that can be gained about rural areas.
While this data leverages permissible WIFI, precise location, advertisements and local content from Facebook user's mobile device, users' raw data are protected using a differential privacy framework that makes an individual point not to be identified from the aggregated data.

A more aggregated application of movement range is the extraction of travel patterns across geographic zones. This data tranche compares inter-country movement from a user's location history, information which is updated daily and available at country admin boundaries. Potential applications include understanding the spread of disease caused by international travel and the economic effect on this for international events.

Colocation data are also included. Colocation in this setting refers to spatial network data that shows the probability of different individual from different location simultaneously being in the same place at a given time. These data are suitable for e.g modeling disease spread by human contact, crises in a given location, and measuring the impact on other human activity. In addition to this, one can produce displacement maps modeling immigration after a crisis, and analyze/predict patterns of populations returning to their originating locations.

Resources:
https://research.facebook.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response/

Colocation Methodology

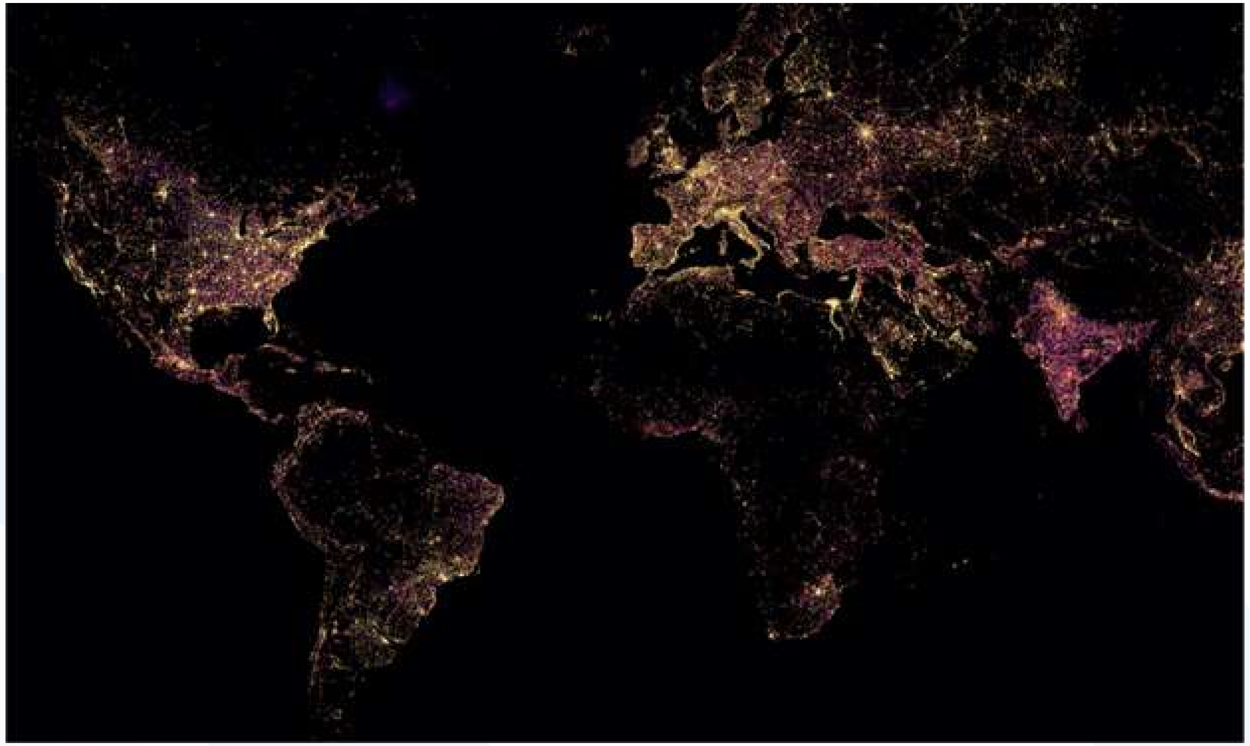
IMAGE SOURCE:
GETTY IMAGE

5.4.4 Extracting maps of regional connectivity/infrastructure

This aspect deals with cellular connectivity and electrical distribution in understanding the existing power sector available in countries at aggregated administrative boundaries.

This data is combination of predictive model, open-sourced surveyed data, remote-sensed satellite imagery and aggregated Facebook users data to develop bespoke solution that can help in planning electrification for vaccine campaign and other immunization development project.

More detail from this data has been employed by the government agency in African countries to create a least-cost technology for energy plans that is sustainable by 2030, which is favorable to the health sector in terms of power supply for critical modern health appliance.



Map 1 leverages the popularity of Facebook to produce global estimates of electrification

Source: dataforgood.facebook.com

Resources:
https://engineering.fb.com/2019/01/25/connectivity/electrical-grid-mapping/

https://impact.economist.com/projects/inclusive-internet-index/downloads/ei-meta_3i_5yr_lookback_report_0.pdf

https://nigeria-iep.sdg7energyplanning.org/

5.4.5 Deriving economic indicators from social media data

Trends regarding relative wealth and business activity within an area can be inferred from users' surveys and posting activities as well as various other actions. Furthermore, the sampling density and frequency of Facebook's 3 billion active users means that these measures can be dynamically updated in time, measuring static economic conditions and how they evolve.

To provide finer-grained, better-localized data in this area, Facebook provides data segmented into commuting zones, showing the geographic areas where people live and work providing a basis for understanding microeconomic patterns at different traditional boundaries. The Facebook-derived indicator, where people spend most of their time during different times of the day can be used to assess the economic viability of a location e.g. residence or business, as well as understand how diseases might be transmitted during a pandemic breakout.

Further providing a basis for building economic indices, the commuting zone data can be used to identify segments of the population different geographical locations based on their movement activity and night-light events. For example, it is possible to estimate that a population consists primarily of an elderly population if that region is residential but exhibits less commuting activity at night in that region.

Facebook's Relative Wealth Index additionally provides details on regional standards of living resolved into a 2.4km x 2.4km grid for more than 100 countries worldwide, with substantial coverage in African countries. This was built using a machine learning model applied to satellite imagery, mobile phone network maps, Facebook's connectivity data, and household surveys.

The Relative Wealth Index can be used to guide interventions and humanitarian aid during pandemic response on targeted populations. This approach was adopted in Nigeria and Togo in providing pandemic relief funds during the COVID-19 pandemic.

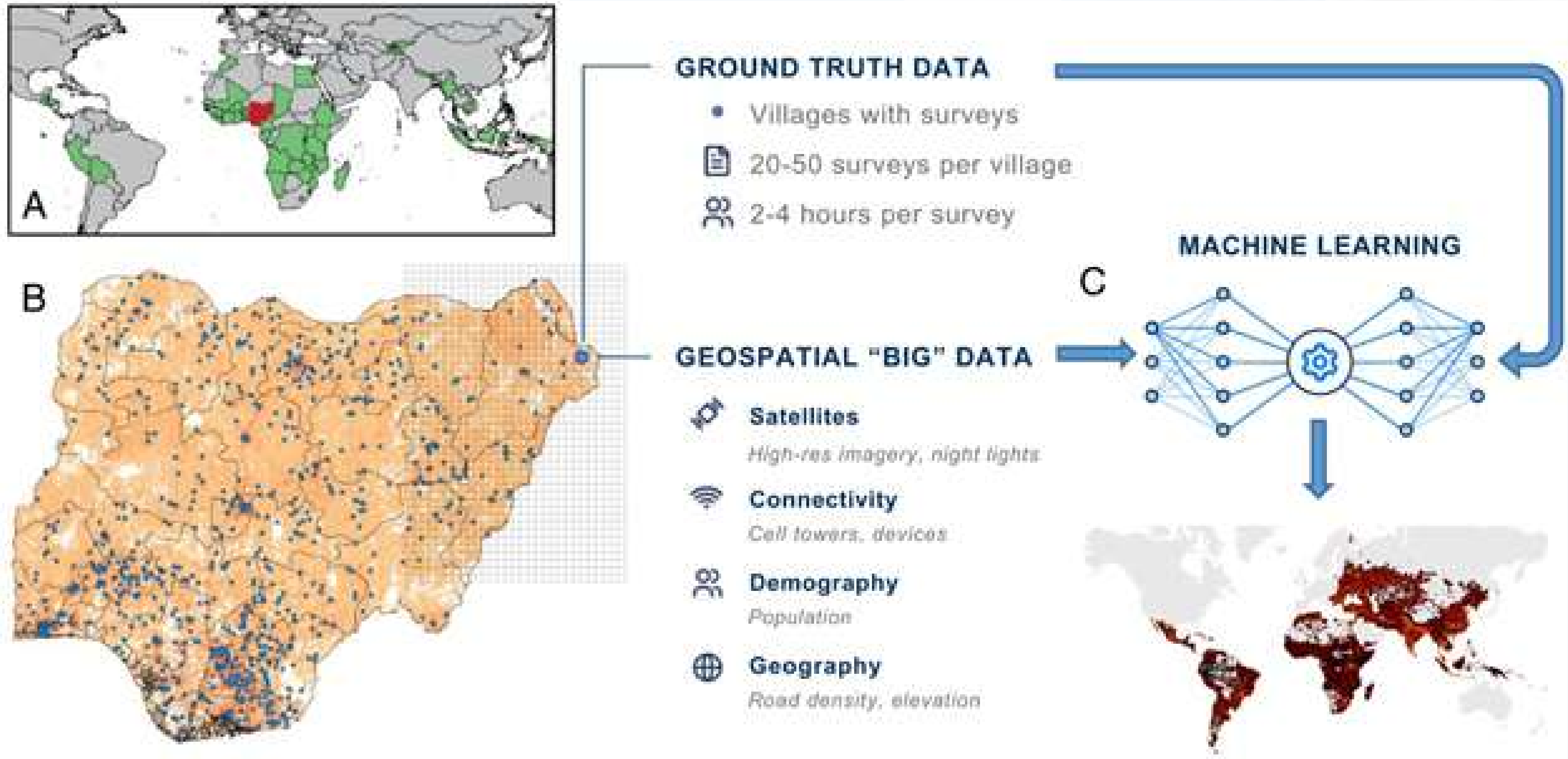


Figure: 1.5
Data flow map for Facebook's estimates of the Relative Wealth Index from 40,860 households in Nigeria

Source:
dataforgood.facebook.com

Resources:
https://dataforgood.facebook.com/dfg/tools/commuting-zones
https://dataforgood.facebook.com/dfg/tools/relative-wealth-index
https://www.nature.com/articles/s41586-022-04484-9
https://www.fastcompany.com/90585079/how-givedirectly-is-finding-the-poorest-people-in-the-world-and-sending-them-cash
https://www.pnas.org/doi/10.1073/pnas.2113658119

5.4.6 Mapping inter/regional social connectivity and social capital

Social connectivity has been shown to be a strong predictor of individual-level factors like mental health and economic prospects. Landmark research by Chetty et. al. (2022, Nature) [19] has demonstrated that socio-economic connectedness as measured using 21 billion Facebook friendships is strongly predictive of economic mobility: establishing friendships and ties across socio-economic classes can drastically improve lifetime socioeconomic outcomes for children who establish such ties, and can be used to motivate and better target policy interventions that aim to foster social connections between low- and high-SES (Socioeconomic Status) individuals and regions.

The Meta *Data For Good* Social Connectedness Index (SCI) measures the strength of connectedness between two geographic areas as represented by Facebook friendship ties. SCI is available at a global scale. Network connection are at international level and can be used as a proxy for unique data form to researchers and non-profit organizations that examines the prevalence and closeness of social connections globally, e.g. an NGO aiming to foster international ties between developed- and developing-countries in order to improve outcomes for the latter.

The SCI metric is developed using de-identified user's activities and information on Facebook which includes location information in the user's profile, as well as device and connectivity detail. Using all these details, a location (i) is assigned to a user, which is subsequently compared against another location(j) using the formula:where,

$$Social\ Connectedness_{i,j} = \frac{FB\_Connections_{i,j}}{FB\_Users_i * FB\_Users_j}$$

\`FB_Usersi\` and \`FB_Usersj\` are the number of Facebook users in locations i and j, and \`FB_Connectionsi,j\` is the number of Facebook friendship connections between the two. These data are available in “Administrative Boundary” 1 and 2 based on the population of each country, with the ISO2 code field denoting the country's code.

Previous work has been done using SCI to understand migration pattern in Europe and can also be leveraged as a basis in disease mapping and disease transfer, e.g. of the flu, at an inter-country level.

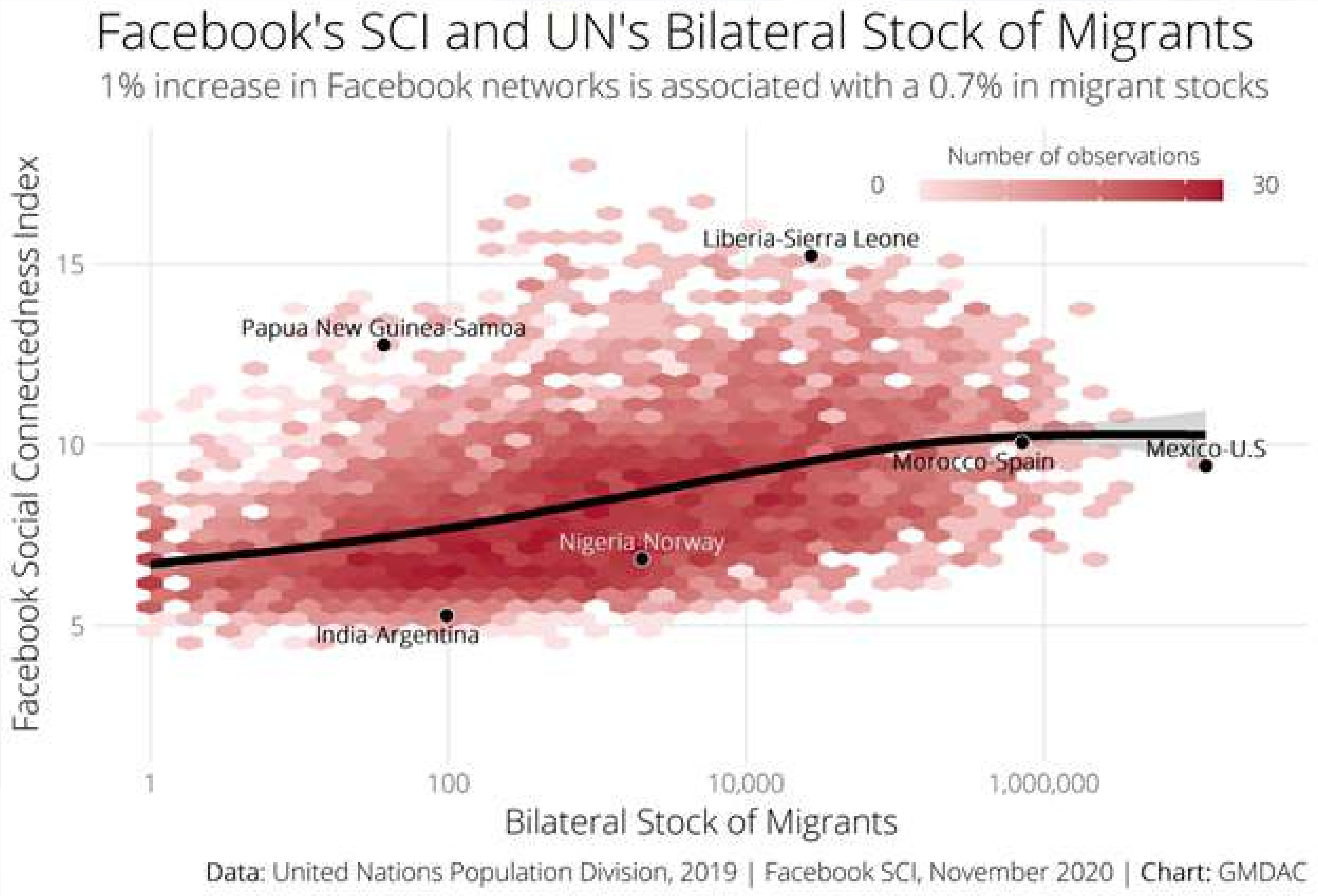


Figure 1.6 shows the level of bilateral connectedness between country-pairs with respect to the migration flow between the pair.

Source:
Resources
https://dataforgood.facebook.com/dfg/tools

6. Applications of Meta *Data for Good* in Health

Over many years of scientific and applied studies, mobility data has proven to be a valuable addition to the process of building real-life solutions to the challenges that affect humans. Especially when triangulated with geospatial data, high-density mobility data exposes the demographic characteristics and spatial positions of points and patterns that can be linked through analysis.

Facebook mobility data can supplement official statistics to monitor changes in mobility patterns in international travel and migration [20]. This change may be driven by local characteristics of the departure country, a pandemic such as COVID-19, or a war such as in Ukraine.

This information is of great importance for NGOs, policy makers, government officials, journalists, and researchers, and has particular relevance and importance in the field of public health, where temporal sampling than is generally available from official statistics is at a premium. As reviewed in the the Impact page of the *Data for Good* website [15], *Data for Good* at Meta have been leveraged in studies providing value in the health sectors .

The impact of the existing and new datasets and visualisations were especially observed to help combat COVID-19 pandemic [21].

6.1 *Data for Good* & Healthcare: An overview

In this section, we offer an overview of how Meta *Data for Good* have been leveraged for health benefits in Africa:

Prior to COVID-19, these data were used by researchers in following areas [20,22]:

- Disaster preparedness projects;
- Vaccination campaigns;
- Deployment during natural disaster times;
- Public health emergency responses with the help of population maps, movement maps and network coverage
- Key health focus areas include analyses of Dengue Fever and Ebola, where mobility data was used to intervene with respect to epidemiological tracking and better targeting delivery of interventions

• Since COVID-19, existing as well as newly developed sets came to rescue to provide dynamic insight into the pandemic's global spread:

- COVID-19 vaccine acceptance: *COVID-19 Trends and Impact Survey*
- COVID-19 community knowledge, practice and behavior: *COVID-19 Preventative Health Survey*
- Community response to mobility policies: *Movement Range Maps*
- COVID-19 cases forecast: *COVID-19 Forecasts*
- *The COVID-19 Map and Dashboard* helped policy makers know whether policies such as social distancing were followed or not and know which approach to invest in to enable them being followed.
- *Symptom surveys were run with around* one million responses a week in the United States. The data helped predict hospital overflow as well as the contamination curve behavior [23].

6.2 Healthcare interventions: In depth

6.2.1 Covid-19 Pandemic: Physical distancing efforts

Around the world, physical distancing policies were communicated in an effort to combat the spread of COVID-19.

The challenge for policy makers, government service providers, and researchers was to know whether the physical distancing measures or stay-at-home orders were indeed followed in a specific area and bearing the predicted benefits flattened disease and hospital usage curves. The Facebook mobility data and maps have been used to provide insight into these important questions [24], yielding further guidance to stakeholders for subsequent messaging, and/or policies.

6.2.2 Increase in maternal healthcare access in Kenya

Around the world, a woman dies every two minutes from preventable causes related to pregnancy and childbirth. Giving birth without the assistance of a skilled provider significantly increases these chances. In Kenya, more than one-third of pregnant women give birth outside of a health facility and without the assistance of a skilled birth attendant[36]. This is caused by various factors such as poverty, inadequate health facilities, and the inability to access emergency health services. According to the Journal of Urban Health[35], increased distance can reduce facility use for reproductive health services including family planning and delivery or child health services. For instance, contraceptive use was significantly lower among women living 2 km or more from a facility compared to women living within 2 km of a facility.

Another study indicated that facility delivery was less likely among women living further than 1 km from a facility compared to women living within 1 km of a facility.

Based on this information, proximity analysis was carried out using Meta movement range data to build a minimum viable product that identifies the closest health facility to a person's location and facilitate the collection and sharing of qualitative data of the health facility based on maternal service delivery. This application would not only assist women to access quality maternal services but also guide stakeholders such as the government in improving and developing health facilities. Link

6.2.3 Emergency care response in conflict/protest areas in Kenya

In Kenya, a student performed an analysis with Meta movement range data to find a pattern between human mobility and conflict, in order to provide quick response to areas prone to conflict and protest. This analysis showed that there is a high level of conflict in medium to high areas of mobility as compared to low-medium areas and also, areas in which people stay put tended to have high level of conflict. The result from the analysis showed that some conflict/protest zones were far away from emergency care sites. This would be instrumental in mapping and planning for a speedy response that could save a lot of lives during a protest and riot. Link

6.2.4 Speedy care for victims of road traffic crashes (RTC) in the Federal Capital Territory (FCT) in Nigeria.

A solution was developed to provide speedy emergency care to victims of road traffic crashes in the federal capital territory of Nigeria, using Meta Movement range data and other georeferenced datasets. These RTC cases range from collision of cars, motorbikes, keke Napep (tricycles), and other vehicles with each other, people and stationary objects. RTCs have become a major public health concern as they are considered; the most common cause of disability, third-leading cause of deaths, and the leading cause of trauma-related deaths. The solution recommended optimal locations to site emergency response teams in order to increase coverage for the RTC sites. Link

6.2.5 Ukraine War: Diaspora mobility in the European Union

On February 24, 2022, Russia invaded Ukraine, causing millions of Ukrainians to flee the country, especially towards the European Union. Whereas data about refugees crossing the borders of neighbouring countries (Hungary, Poland, Slovakia) were carefully collected by local and international organizations, it was much challenging to receive information of refugees migrating to other countries of the EU due to the free movement agreement in the Schengen Area. Ensuring adequate provision of health, housing, economic, and other administrative resources to displaced Ukrainians requires detailed dynamic mapping and prediction of migration flows in response to the war.

Drawing upon the *Data For Good* data suite, researchers[26] came to the rescue to estimate the number of Ukrainian migrants in the different EU countries with regional-level resolution, defined by the NUTS-3 (Nomenclature of Territorial Units for Statistics-level 3) demarcations. With the help of two of its datasets—the Monthly Active Facebook Ukrainian speaking users (MAUs) and the Facebook Social Connectedness Index (SCI) measuring social connectedness between different geographical locations—researchers found a number of results:

1. MAUs provide insight about the increase of Ukrainian diaspora in the different EU countries as a consequence of the Ukrainian war.
2. SCI assesses that social connectedness as driver of immigration is stronger in some countries (e.g. Czech Republic) and weaker in others (e.g. Germany).

7.

Data Triangulation and Use Cases of Meta *Data for Good*

7.1 Mobility Data Taxonomy

Mobility data can be presented in the form of GPS traces, mobile phone traces, vehicles mobility traces, etc. This vast and diverse panel of data sources, collection types and technology categories, although valuable, make it cumbersome to be leveraged for researchers and decision makers, hence the need for classification.

Mukhopadhyay King, Nawab, and Obraczka [34] suggest a taxonomy to classify open-source mobility traces based on mobility mode and data source on one side and based on information category on the other side. They identify Pedestrian and Vehicular as mobility modes. The data sources cover both the collection infrastructures (laptops, smart vehicles, etc.) and measurement medium (oximeter, WIFI, camera, etc.).

Finally, the information category is further categorized into Connectivity (e.g., network speed), Location (e.g., distance travelled), Health (e.g., body temperature) and Lifestyle (e.g. power usage).

7.2 Metrics: Transforming mobility data into insight

Facebook provides a wealth of data on individual-level mobile device usage, location from network access points, access to electricity and the means of digital communication, balance of social/business activity, and individual social activity and quantity of connections, comprising billions to trillions of heterogeneous datapoints. In order to transform these into insight, however, the data must be structured, aggregated, and modeled. In this section, we review how, after acquiring human mobility data and identifying what types of data we need, we can gain new insights and understanding.

Analysis of this sort must always be viewed in the light of the socio-economic context tied to human mobility. Meredith et al.[27] encapsulates this point well: "Human mobility patterns are a reflection of behaviors, ranging from routine (e.g., commuting daily for work or school, traveling for holidays and religious gatherings, or seeking seasonal work opportunities) to irregular. (e.g., relocating due to environmental changes or crises or social distancing due to a pandemic)" In fact, the benefit of human mobility data lies in the understanding of underlying human behavior which directly impacts everyday life in areas such as natural resources management, infectious disease transmission, urban planning, etc. Some metrics for either individual or community movement pattern have been set up to characterize human mobility, such as:

• *The traveling distance or jump length:* the distance that separates two consecutive points visited by a person. This gives insight to an urban planner for example about what ranges of distances people are willing to walk and plan the city accordingly in order to reduce vehicle's pollution.

• *The radius of gyration* indicates the spread of locations visited by a person. For example, Figure 1.1 shows that the first person has visited closer locations, whereas the person B's locations are more spatially spread. This may depict two different categories of people that work closer to home versus those who commute longer distances.

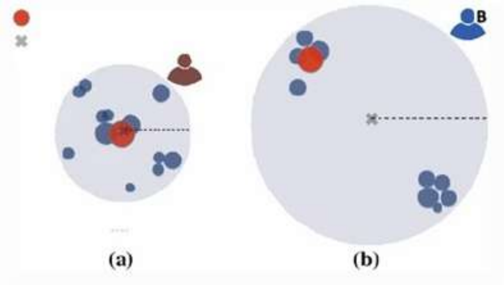


Figure 2.1: The radius of gyration is shown by the dotted line (Pappalardo et al.[28])

• The mobility entropy indicates the connections between different locations visited. Figure 2.1 shows that Person X has visited mostly three locations whereas Person Y has visited many more locations with similar frequency. Person X's behavior may show a person whose locations are mainly home, office, and children's school, for example, whereas Person B may work in door-to-door sales, marketing or delivery.

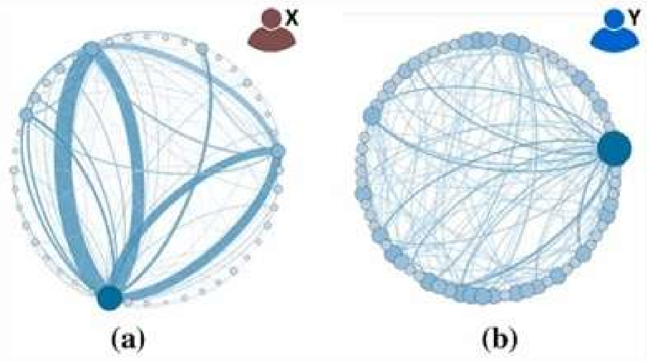


Figure 2.2: The radius of gyration is shown by the dotted line (Pappalardo et al.[9])

• *The origin-destination matrix* (OD matrix) is a key tool for transportation engineers. It aggregates individual number of trips from a given origin to a given destination based on either a specific time or specific routine such as home-work commute. Thanks to mobility data, the OD matrix can be dynamic in time and built to represent a higher temporal resolution.

• *Mobility motifs* are networks frequently travelled by a person. More than point locations, networks reflect human activities and their connections. These are especially interesting knowing that they are common to various people regardless of where they live. Most human beings regardless of where they live share activities such as sleeping at home, going to work, going to shopping, etc. Thus, Figure 2.3 show similar motifs for two different persons. This kind of information can be crucial for urban planners or epidemiological models.

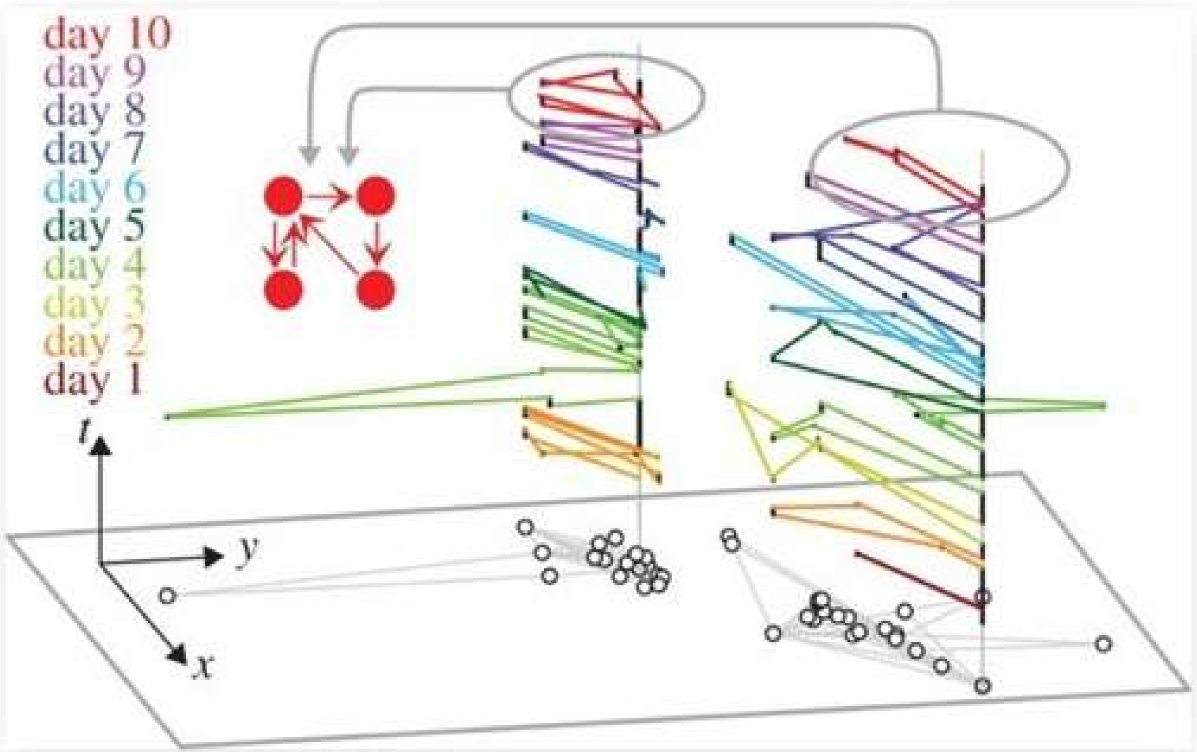


Figure 2.3: Mobility Motifs (Schneider et al[10])

7.3 Triangulated GIS data: drawing from diverse data formats

Data triangulation is a technique used in research to increase the validity of findings by collecting data from multiple sources or using multiple methods. The idea is that by gathering data from different perspectives or using different methods, or by comparing the results from each source, researchers can increase their confidence in the validity of their findings.

Data triangulation is often used in qualitative research, but it can also be used in quantitative research, and is of particular use in sensitive topic areas where the data can be easily manipulated. Just as a lawyer might build a case by drawing patterns from physical and circumstantial evidence, as well as establishing motive, data scientists can seek to cross-reference their findings across multiple data sources so as to reduce bias and increase confidence in the findings.

We will examine below some examples of data sources and the advantages they provide when triangulated with the META's privacy-preserving mobility data. However, it is important to note that combining data from different sources can be complex and requires careful consideration of potential sources of bias or inaccuracies in the data. It's also important to check and make sure that the data is collected on similar time ranges and the same geography so that it can be effectively compared.

Overall, the ability to triangulate data from different sources can provide a more comprehensive and accurate understanding of people's movements and behavior, which can be useful for a variety of purposes, including urban planning, public health research, and marketing.

Meanwhile, the benefit and actual use cases will depend on the quality and granularity of data that can be collected and analyzed. Finally, it is important to keep in mind of data privacy and data protection issues that may arise while gathering and using the data.

7.3.1 Deeper measures of economic development

One indicator of a region's level of economic development is not just binary access to the internet, but also the quality of a user's internet connection. Internet speed test data from more that 18 million contributors around the world can help decision-makers make informed decisions around internet connectivity, policy, development, education, disaster response, public health, and economic growth[37].

Combining data from META's mobility reports with Internet speed test data could be useful for gaining a more complete understanding of how people are using the internet in a specific location.

META's mobility data provide information about the extent to which people use different place-categories, such as parks, businesses, and residential areas, as well as how this usage has changed over time. Internet speed test data, on the other hand, provides information about internet connection speeds and quality.

By combining these two data sets, researchers could get a better sense of how internet usage patterns are related to internet connectivity. For example, an investigator could see if there are specific times or locations where internet usage is higher or lower, and if this corresponds to changes in internet speed or quality.

Use Cases

Network Planning and Optimization: by understanding usage patterns and internet quality in different locations, a telecom company could identify and address bottlenecks and improve its service provision.

Urban Planning: by understanding the relationship between internet usage patterns and physical locations, a city planner could identify areas where better connectivity could improve quality of life and use of public spaces.

Market Analysis: by understanding how internet usage patterns vary across different types of businesses or neighbourhoods, a company could identify potential markets for its products or services.

7.3.2 Satellite imagery and night-time luminosity datasets

Triangulating META's mobility data with satellite imagery and nighttime luminosity data can provide a more complete picture of how people are moving and where they are spending their time. This information can be useful for a number of different use cases.

Whereas META's mobility data provides spatially and temporally resolved data on person locations, satellite imagery, provides a detailed visual representation of an area's static built environment, such as buildings, roads, and open spaces.

Use Cases

· **Urban Planning:** Triangulating this data can help planners understand how land use patterns are changing and make more informed decisions about transportation infrastructure and the location of public amenities.

· **Disaster Response:** Combining mobility data with satellite imagery can help emergency responders understand the extent of damage from a natural disaster and make more informed decisions about where to deploy resources.

· **Economic Development:** The analysis of nightlight data in conjunction with mobility data can help governments and businesses understand economic activity in a given area, which can be useful for making investment decisions.

· **Public Health:** Analyzing mobility data in combination with satellite imagery and nightlight data, researchers can gain a more complete understanding of how social distancing measures are affecting people's movements and identify potential hotspots for disease transmission.

· **Crime Analysis:** This level of data triangulation can be useful to understand patterns of criminal activity, such as where burglaries occur in relation to a population density, or how nightlight activity may relate to crime occurrence.

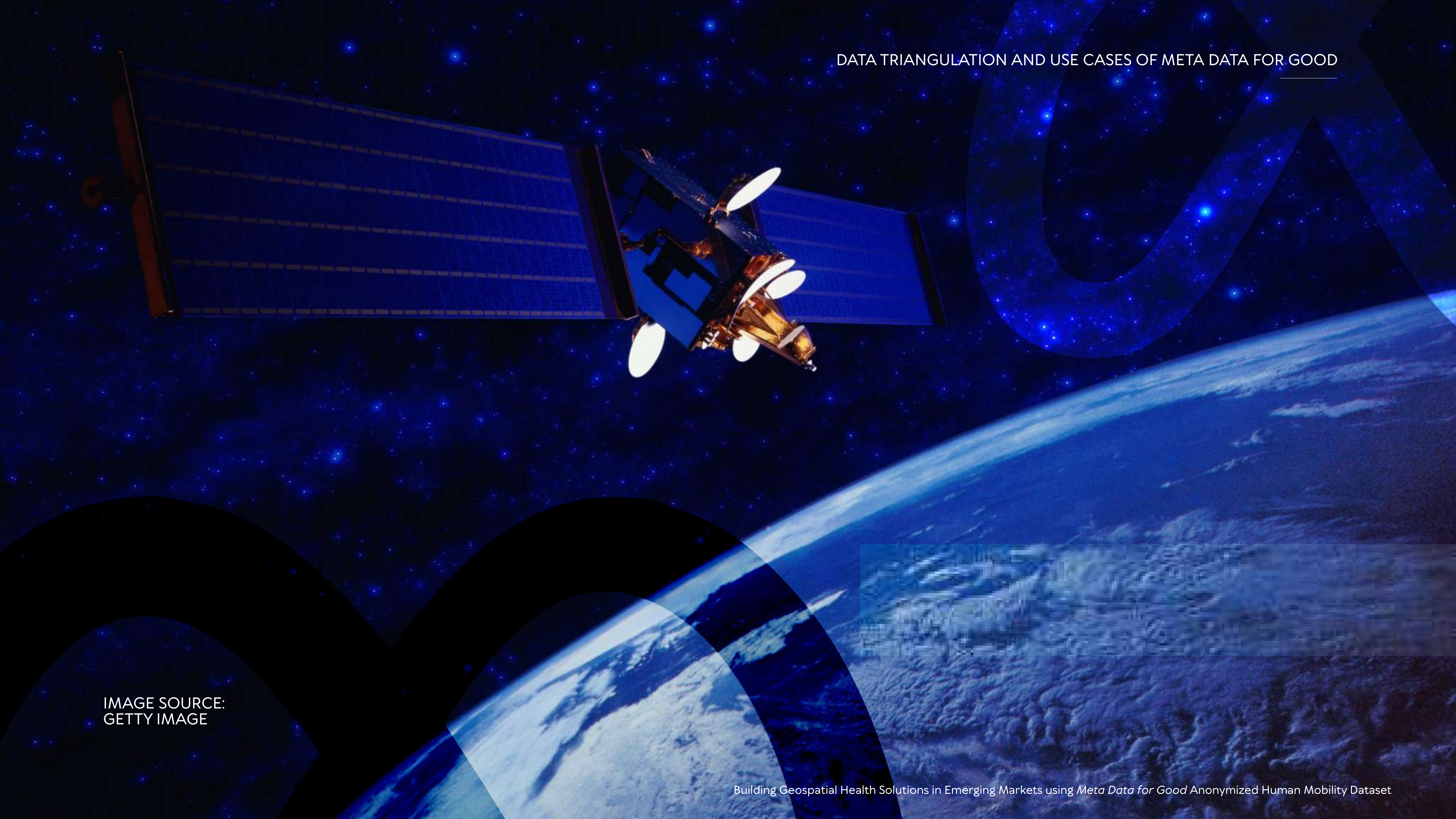


IMAGE SOURCE:
GETTY IMAGE

7.3.3 Household profiling data

Triangulating household profiling based on satellite imagery data with Facebook mobility data can provide a more detailed understanding of people's movement and where they are spending their time. This information can be useful for a number of different use cases, including:

Use Cases

· **Targeted Marketing:** Combining mobility data with household profiling can help businesses identify areas where there is high foot traffic and target their advertising and promotions to those areas.

· **Social Services:** Analyzing household profiling data in conjunction with mobility data can help governments and non-profits identify areas where there is a high concentration of low-income households, which can be useful for providing targeted social services.

7.3.4 Settlement and farmland mapping

Extrapolating Facebook mobility data using settlement and farmland datasets can provide important insights into the relationship between population movement and land use. For example, by comparing Facebook mobility data with settlement and farmland datasets, we will be able to identify patterns of migration and urbanization, as well as changes in agricultural land use.

This type of data triangulation can have a wide range of use cases. For example, it could be used to inform urban planning and development decisions, as well as agricultural policy and land management. Additionally, it could be used to monitor and respond to natural disasters, such as floods or droughts, as well as to track the spread of infectious diseases.

Other potential use cases include monitoring climate change, analyzing economic activity, and studying the impact of transportation infrastructure on population movement.

7.3.5 Triangulating social media activity with hotspot access

The nature of a user's internet connection (mobile network vs. public hotspot) during a network access event can be an indicator of the type of activity they are engaged in. Facilities with built-in WiFi access points are likely to be schools, health facilities, or market locations, and social media use within such environments will likely differ from activity while commuting.

Triangulating Facebook mobility data with hotspot datasets can provide valuable insights into how people are utilizing public amenities, services, and facilities.

By comparing Facebook mobility data with information on schools, health facilities, and market locations, researchers can gain a better understanding of attendance and usage patterns, as well as the impact of various factors such as school closures, health interventions, and natural disasters on population movement.

Use Cases

- Informing education policy by identifying patterns of school attendance, dropout rates and the impact of school closures on population movement.

- Improving healthcare by understanding patterns of health-seeking behavior and the impact of health interventions on population movement.

- Monitoring and responding to public health crises by identifying patterns of population movement and usage of health facilities during outbreaks of infectious diseases.

- Analyzing economic activity by identifying patterns of population movement and usage of market locations, transportation infrastructure, different areas of a city, or critical facilities during natural disasters.

7.3.6 Population demographic data (e.g. age/gender distributions)

Triangulating Facebook mobility data with population data can provide valuable insights into the demographic characteristics of individuals who residing or moving around in a specific area. For example, by comparing the age and gender distribution of individuals who are actively using Facebook with population data, researchers can gain a better understanding of how different demographic groups are interacting with technology.

Use Cases

· **Public Health:** By analyzing the mobility data of individuals in a specific area, researchers can identify patterns in population movement that may be associated with certain health outcomes. This could be useful for identifying areas where there may be a higher risk of infectious disease transmission, or estimating regional rates of routine exercise.

· **Marketing:** By understanding the demographic characteristics of individuals who are moving around in a specific area, businesses can target their marketing efforts more effectively. For example, a business may want to target its advertising to individuals who are most likely to be interested in its products or services.

· **Urban Planning:** By analyzing population movement data, urban planners can gain a better understanding of how different demographic groups are using different areas of a city. This can be useful for identifying areas that may need additional infrastructure or services.

· **Transport Planning:** By analyzing population movement data, transport planners can gain a better understanding of how different demographic groups are using different forms of transportation. This can be useful for identifying areas where there may be a need for additional public transport services or bike lanes, for example.

· **Security:** By analyzing population movement data, public safety officials can gain a better understanding of how different demographic groups are using different areas of a city. This can be useful for identifying areas where there may be a higher risk of crime or other public safety concerns.

7.4 Using anonymized and aggregated mobility data: ensuring privacy and verifying representativeness

Two main challenges are noted in the use of human mobility data: privacy and representativeness[3].

7.4.1 The Privacy challenge

With any kind of human-related data, one parameter of paramount importance has to be considered: **Privacy**. Human mobility data science is not about tracking or spying on people but about helping the same people make strategic decisions for their work, health, business.

In fact, the sensitive nature of the data carries the risk of causing security concerns or being weaponized as an instrument of control that undermines basic human rights such as freedom of opinion.

Proposed Solution:

• *Anonymization of data* involves the disabling of the identifiability of personal information and data. Currently, different methods are used, such as rejecting personal data and introducing some noise to acceptable limits[4].

• *Aggregation of data* by time or space. It makes it more difficult, if not impossible, to identify personal information of an individual from the information of thousands of people who enter or leave from a city A to a city B every week[5].

7.3.2 The representativeness challenge

In order to get truly anonymized and aggregated data, it is necessary to have access to sufficient data within a specific area and time.

Therefore, the results of these methods are dependent on the size of the data. This infers an inherent bias which results from the unequal access to mobile phone on one side and unequal usage of mobile phone on the other side. Nevertheless, when using data for decision making, we should ensure that these data statistically represent the "population of interest".

Proposed Solution:
Solutions include extrapolation of data i.e., adjusting the data according to observed mobile phones market share [5] or de-biasing procedures such as the one developed by Schlosser et al. [6].

8.

The Africa Geoportal

— Geospatial Tools, Data, and Trainings

Africa has seen a flourishing of data science and data-driven economic activity in recent years. A particularly valuable addition to this vibrant ecosystem is the Africa GeoPortal, a tremendous resource for researchers seeking to understand a variety of aspects of African geography, sociology, demographics, economics, and public health.

When working with mobility and geospatial dataset, it is important to use a platform that allows flexibility to work with vector, raster and time series datasets. Africa GeoPortal is a platform that allows professionals across Africa to triangulate data from various sources as well as use tools and applications that leverage ArcGIS technology. Africa GeoPortal is a platform that combines data sourced from various entities, mostly authoritative data sources, together with tools that support online visualization and analysis and learning resources that can be used by professionals to build their personal technical capacity and learn how to work with spatial data. AfricaGeoPortal has different datasets —available at the country level—that meet a range of data needs. The portal features strong use cases across different sectors, useful in generate ideas and building expertise, empowering individuals, NGOs, and businesses to support communities of users across Africa. The platform's dashboards, data, geospatial tools and learning resources are tailored to Africa-centric use cases, allowing users at various stages of their data science journey (student, practitioner, expert, translational policymaker) to take advantage of when performing their analysis.

This platform is suitable for all classes of users, from experts to beginners who have low coding skill. The platform works online and so provides cloud space for analysing, visualizing, collaborating and storing your works for later usage. The systerm's python notebook interface also combines drag-and-drop features integrated with code, and users can select the data interaction format most suitable to them.

9. Bibliography

1. ESRI. History of GIS. url: https://www.esri.com/en-us/what-isgis/historyof-gis. (accessed: 21.01.2023).

2. Mohamed Mokbel et al. "Mobility data science (dagstuhl seminar 22021)". In: Dagstuhl reports. Vol. 12. 1. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2022.

3. Maryam Kiashemshaki, Zhiren Huang, and Jari Saramäki. "Mobility signatures: a tool for characterizing cities using intercity mobility flows". In: Frontiers in big Data (2022), p. 18.

4. Daniel T Citron et al. "Activity Space Maps: a novel human mobility data set for quantifying time spent at risk". In: medRxiv (2021).

5. Kamil Smolak, Witold Rohm, and Katarzyna Sila-Nowicka. "Explaining human mobility predictions through a pattern matching algorithm". In: EPJ Data Science 11.1 (2022), p. 45.

6. Chaoming Song et al. "Limits of predictability in human mobility". In: Science 327.5968 (2010), pp. 1018–1021.

7. Andrea Cuttone, Sune Lehmann, and Marta C González. "Understanding predictability and exploration in human mobility". In: EPJ Data Science 7 (2018), pp. 1–17.

8. Douglas Do Couto Teixeira et al. "The impact of stationarity, regularity, and context on the predictability of individual human mobility". In: ACM Transactions on Spatial Algorithms and Systems 7.4 (2021), pp. 1–24.

9. Douglas do Couto Teixeira, Jussara M Almeida, and Aline Carneiro Viana. "On estimating the predictability of human mobility: the role of routine". In: EPJ Data Science 10.1 (2021), p. 49.

10. Yuan Liao, Sonia Yeh, and Gustavo S Jeuken. "From individual to collective behaviours: exploring population heterogeneity of human mobility based on social media data". In: EPJ Data Science 8.1 (2019), pp. 1–22.

11. Satchit Balsari et al. "Human Mobility Data in Public Health Emergencies". In: (2022). url: https://www.crisisready.io/wp-content/uploads/2022/06/The-Use-ofHuman-Mobility-Data-in-Public-Health-Emergencies.pdf.

12. Aidan Findlater and Isaac I Bogoch. "Human mobility and the global spread of infectious diseases: a focus on air travel". In: Trends in parasitology 34.9 (2018), pp. 772–783.

13. Robert Moss et al. "What can urban mobility data reveal about the spatial distribution of infection in a single city?" In: BMC public health 19.1 (2019), pp. 1–16.

14. Frank Schlosser and Dirk Brockmann. "Finding disease outbreak locations from human mobility data". In: EPJ data science 10.1 (2021), p. 52.

15. Meta. *Data for Good*. url: https://dataforgood.facebook.com/. (accessed: 23.01.2023)

16.. Centre for humdata. Community Q&A: Data For Good At Meta. url: https://centre.humdata.org/community-qa-data-for-good-at-meta/. (accessed: 23.01.2023).

17. DEMED Research Project- University of Glasgow. Webinar on Facebook as a Research Tool: Session 1 (Spring 2022) - Academic Partnerships at Meta. url: https://youtu.be/oGfPkyfXIQU. (accessed: 23.01.2023).

18. Data Science Campus. Using Facebook data to understand changing mobility patterns. url: https://datasciencecampus.ons.gov.uk/using-facebook-data-to-understandchanging-mobility-patterns/. (accessed: 23.01.2023).

19. Raj Chetty et. al. "Social capital I:measurement and associations with economic mobility". In: Nature. Vol. 608, pg. 108-121. 2022.

20. Migration Data Portal. Quantifying international human mobility patterns using Facebook network data. url: https://www.migrationdataportal.org/data-innovation/ Quantifying-international-human-mobility-patterns-Facebook-network. (accessed: 23.01.2023).

21. Meta. *Data for Good*: New Tools to Help Health Researchers Track and Combat COVID-19. url: https://about.fb.com/news/2020/04/data-for-good/. (accessed: 23.01.2023).

22. London School of Hygiene & Tropical Medicine (LSHTM). Facebook Disease Prevention Maps developed to help respond to outbreaks. url: https://www.lshtm.ac.uk/ newsevents/news/2019/facebook-disease-prevention-maps-developed-helprespond-outbreaks. (accessed: 23.01.2023).

23. The Washington Post. Mark Zuckerberg: How data can aid the fight against covid-19. url: https://www.washingtonpost.com/opinions/2020/04/20/how-data-can-aidfight-against-covid-19/. (accessed: 23.01.2023).

24. Crisis Ready. Facebook Mobility Map. url: https://www.crisisready.io/resources/ facebook-mobility-map/. (accessed: 23.01.2023).

25. Escamilla, V., Calhoun, L., Winston, J. et al. The Role of Distance and Quality on Facility Selection for Maternal and Child Health Services in Urban Kenya. J Urban Health 95, 1–12 (2018). https://doi.org/10.1007/s11524-017-0212-8

26. Umberto Minora et al. "Innovative human mobility data for fast crisis response". In: (2022).

27. Hannah R Meredith et al. "Characterizing human mobility patterns in rural settings of sub-Saharan Africa". In: Elife 10 (2021), e68441.

28. Luca Pappalardo et al. “An analytical framework to nowcast well-being using mobile phone data”. In: International Journal of Data Science and Analytics 2.1 (2016), pp. 75– 92.

29. Christian M Schneider et al. “Unravelling daily human mobility motifs”. In: Journal of The Royal Society Interface 10.84 (2013), p. 20130246.

30. Precisely & Unacast. Human Mobility in an Evolving World: Why understanding populations and the way they move has become critical. url: https://www.precisely.com/ resource-center/ebooks/human-mobility-data-understanding-populations-and-the-way-they-move. (accessed: 21.01.2023).

31. University. Anonymizing Mobility Data for Civic Use Cases. url: https://univercity.ai/mobility-data-anonymization-and-aggregation/. (accessed: 21.01.2023).

32. Kristofer ˚Agren, Pär Bjelkmar, and Elin Allison. “The use of anonymized and aggregated telecom mobility data by a public health agency during the COVID-19 pandemic: Learnings from both the operator and agency perspective”. In: Data & Policy 3 (2021).

33. Frank Schlosser et al. “Biases in human mobility data impact epidemic modeling”. In: arXiv preprint arXiv:2112.12521 (2021).

34. Sinjoni Mukhopadhyay King, Faisal Nawab, and Katia Obraczka. “A Survey of Open Source User Activity Traces with Applications to User Mobility Characterization and Modeling”. In: arXiv e-prints (2021), arXiv–2110.

35. Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018a. “Social connectedness: Measurement, determinants, and effects.” Journal of Economic Perspectives, 32(3): 25980.

36. “Free Maternal Healthcare in Kenya Saves Live.” YouTube video, 5:15. Posted by Health Policy Plus, July 22, 2020. Available at: https://youtu.be/5wjaLh3CQL8.

37. https://www.ookla.com/ookla-for-good/open-data

10.

Appendix:

Additional resources

A major policy goal of the data science ecosystem in Africa is to make the tools, concepts, and methods of data analysis accessible to a range of users on the continent. Here, we provide links to 9 learning videos in English and French, as well as learning content and slides from introductory data science classes covering topics in data science methodology. These resources are useful for anyone seeking to wet their feet or acquire greater expertise in the analysis of mobility data.

A.1 Video tutorials to spatial analysis tools and concepts

1. Introduction to GIS and Spatial Analysis

This training's objective is to enable learners to understand the basic concepts of Geospatial Information Systems (GIS) and how spatial analysis is conducted.

The Introductory session covers the areas like the GIS Concepts, Applications of GIS, Types of Spatial Data, Spatial Tools, and a hands-on session.

The hands-on session exposed learners to the Africa Geoportal platform of the ESRI and how to carry out spatial analysis in a simple step-by-step and drag and drop approach.

Visit this link to learn more
(https://youtu.be/qSOMn-ZIStw)

2. Introduction to GIS and Spatial Analysis using ArcGIS Online (Vector Analysis)

This training exposes learners to spatial analysis in the ArcGIS online interface. Learners will learn the types of Vector Analysis and the different tools that can be used to achieve these types of analysis. The training explores a use case to look at the areas with high number of conflicts in Kenya and then proceed to access the proximity of health facilities to conflict prone locations.

Visit this link to learn more:
(https://youtu.be/jUIlp4MeMxg)

3. Vector Analysis with Python

This training was a continuation of the previous training. It focuses on the use case of Conflict data and health facility locations. The training utilizes the buffer analysis tool to understand the number of conflict locations that fall within a 1km radius of the nearest health facility locations, exploring data in the attribute table and line distance between the conflict zones and the nearest health facilities in KM.

Visit this link to learn more:
(https://youtu.be/Rav6KSS9b_4)

4. Introduction to Time Series Using ArcGIS

This training session introduces participants to Facebook Movement Range dataset, how to prepare their python environment for the geospatial analysis, how to add data for the analysis, how to query data to the specific sample of the dataset and how to work with datetime dataset.

Visit this link to learn more:
(https://youtu.be/LHvTNSj23Gg)

5. Introduction to ArcGIS Pro (Vector and Raster Analysis)

This training focuses on the introduction of the ArcGIS Pro, the key terms in the ArcGIS Pro, the Capabilities of ArcGIS Pro, The ArcGIS project, and a short hands-on session to deepen the understanding of participants on how to visualise their data, style map on ArcGIS, how to use expressions or SQL to manipulate their datasets and how to share map.

Visit this link to learn more:
(https://youtu.be/iEOa_dJze88)

6. Data Processing in ArcGIS Pro - (Vector and Raster Analysis)

This training covers how to generate Tessellation, work with Raster data, understand the Population density in its constrained and unconstrained formats, and process health facility data. The training concluded with a hands-on session.

Visit this link to learn more:
(https://youtu.be/5AR_8eJ7JEk)

7. Creating Instant App on ArcGIS

This training introduces participants to Instant App and the step by step of how to create Instant App. The training also include how to utilize existing templates to speed up the creation of Instant App

Visit this link to learn more:
https://youtu.be/opUK9eqO2Zs

8. Sharing Maps and other Solutions on ArcGIS (Case Study)

This training explores the case study of School at the Risk of Health challenges due to their distance to the nearest health facility. It explores the combination of Health Facility Data and School Dataset. It also shows how to publish the solution developed to allow external users to be able to access the solution and make informed decision based on the insights from the data.

Visit this link to learn more:
https://youtu.be/fSGxH8zI9Ac

9. Aggregating and Sharing Solutions on ArcGIS

This session of the training builds on the training on Creating Instant Apps. It walks learners through how to set up the Story Map on the ArcGIS. This training also helps learners to understand the concept of data triangulation with a use case in identifying the next best location to build the next health facility.

Visit this link to learn more:
https://youtu.be/elxWff6We1U

11.

Additional resources

Learning geospatial analysis for mobility data can be daunting and difficult at times, in part due to the complexity of the concepts. In addition, many potential users may not be motivated to gain expertise in this area because they have not yet been exposed to to application domains uniquely geared towards the African context. Another factor is the underdevelopment of quality open-source datasets that would allow for a more active ecosystem of data learners and users with which to prototype their own experiments and analyses.

At the present time, many of these issues are being eradicated with the rapid emergence of innovative data sources from users's constant internet activities, IOT sensors, and the easy availability of satellite imagery. Given the availability f these resources, now more than ever there is a greater need for organizations to build capacity in deriving insight from this data. Here are some additional resources users can use to learn how to work with spatial data.

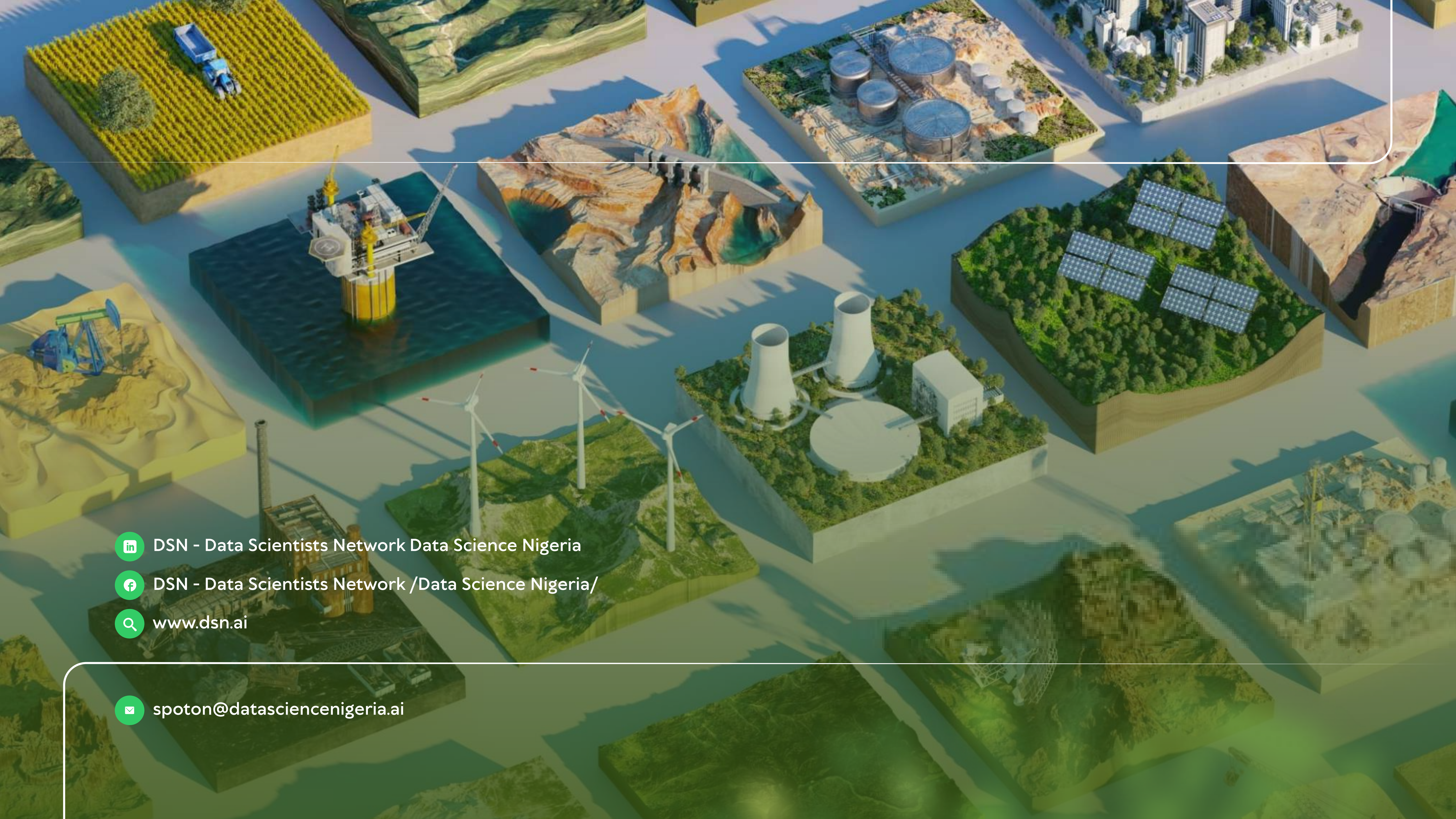
· Introduction to Spatial Analysis
Gentle Introduction to GIS - QGIS

Introduction to Geographic Information Systems - Kang-Tsung Chang
– Geospatial Analysis A Comprehensive Guide to principal Technical and Software Tools.
- Michael J Smith
Spatial data science for sustainable development

Mobility Data Analysis
Mobility Data Seminar
Report - Dagstuhl 2021
Scikit-Mobility Tutorials
Moving Pandas Tutorials
A Survey on Deep Learning for Human Mobility

· Spatial Analysis with ArcGIS/QGIS
Introduction to ArcGIS Online
QGIS Tutorials and Tips

·Spatial Analysis with Python
Automating GIS Processes 2022 - Helinski
Python Foundation for Spatial Analysis (Full Course Material) - Ujaval Gandhi
Spatial Data Programming with Python
- Michael Dorman



DSN - Data Scientists Network Data Science Nigeria

DSN - Data Scientists Network /Data Science Nigeria/

www.dsn.ai

spoton@datasciencenigeria.ai