Photo by Kurt Cotoaga on Unsplash

Use of mobile phone data for enhancing migration statistics
The Gambia use case

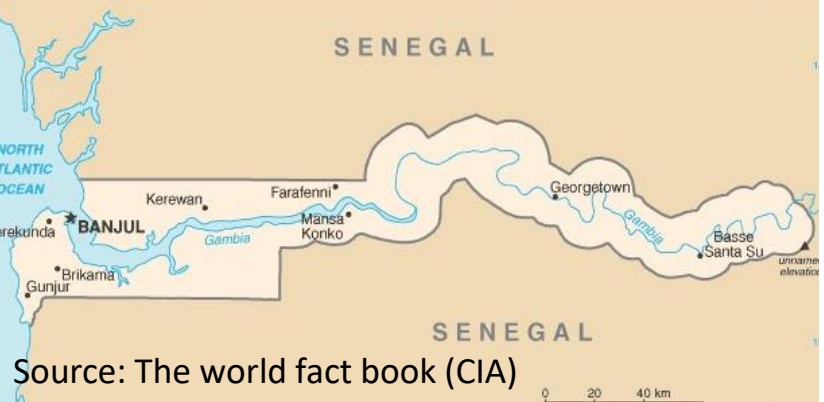
Presented by: Lamin L Dibba

Affiliation: Gambia Bureau of Statistics

Overview

1. Context, engagement, and data access
2. Selected results from internal migration use case
3. Adjusting representative biases using data from Census

Country context: *The Gambia*



Source: The world fact book (CIA)

- **Large spatial disparities** with the country and economic activities
- **Legacy of high migration** offering access to opportunities but also source of external vulnerability
- Dependent on domestic/international remittances from **migrants**

Migration statistics is crucial for designing policies in The Gambia.

Early engagement:

Before Covid-19

- Supported the GBoS (NSO) in complementing migration statistics.
- Established a platform to discuss the policy relevance.
- Adopted technical and organizational measures to protect privacy.

Needs: *after Covid-19 onset*

- An evidence base for examining the impact of COVID-19 and enforcement measures.
- Timely and quality data to assist NSO's data collection – phone survey
- Technical support in producing more statistics in addition to migration statistics.

GBoS THE GAMBIA BUREAU OF STATISTICS
PURA Public Utilities Regulatory Authority Equity in Development
WORLD BANK GROUP
THE UNIVERSITY OF TOKYO

Stakeholders

Data access PURA brought the regulatory mandate and technical capacity.

Method GBoS, The World Bank, and The University of Tokyo (Spatial Data Commons) guided and motivated the analytics.

Data protection Policy relevance and data privacy was assessed based on the legal framework.

GBoS THE GAMBIA BUREAU OF STATISTICS

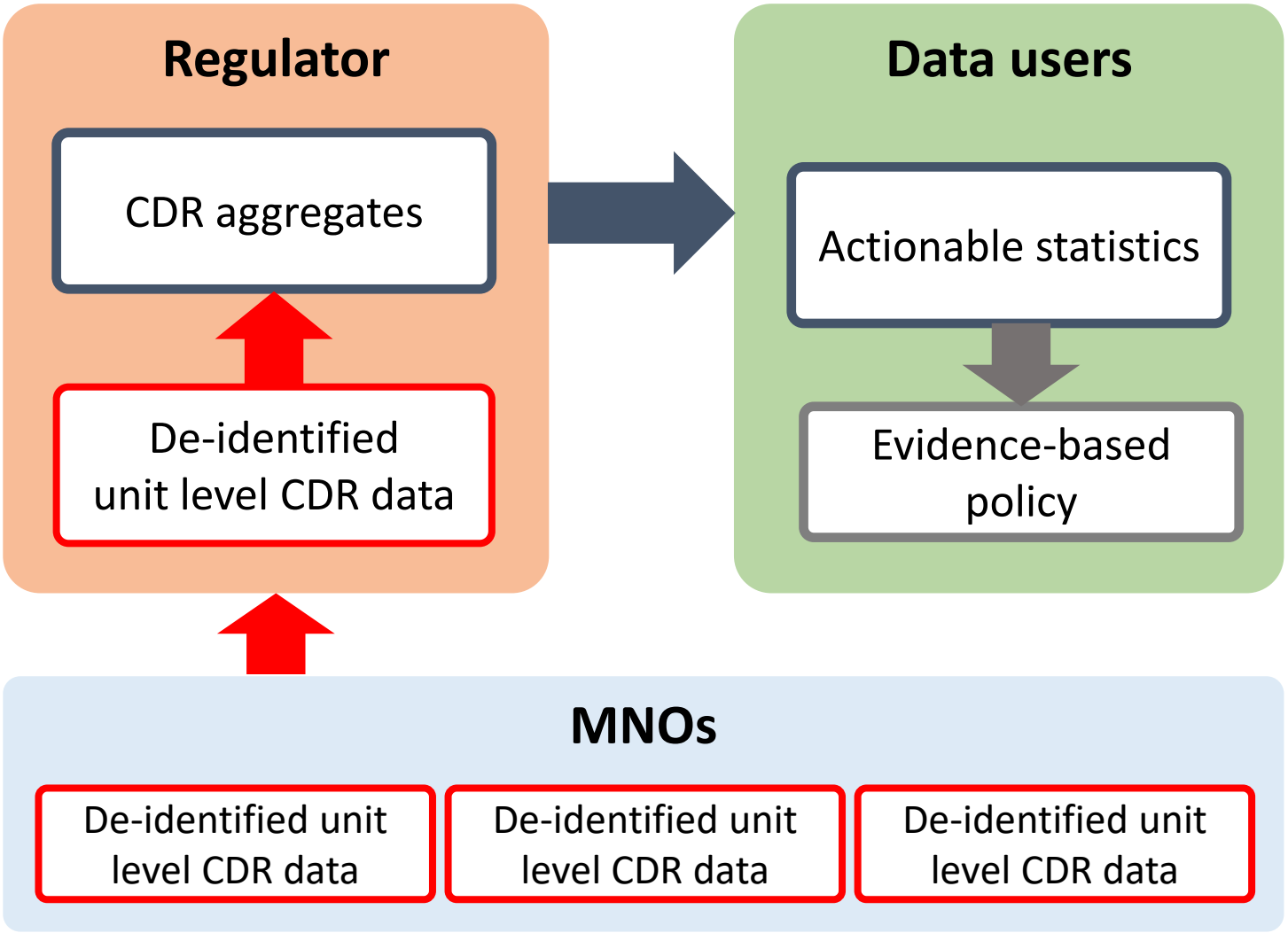
PURA Public Utilities Regulatory Authority Equity in Development

WORLD BANK GROUP

THE UNIVERSITY OF TOKYO

Data access model to protect data

- **Unit level data stay on regulator's premise**
- **CDR data are processed on the PURA's premise using the hardware provided by The World Bank**



Privacy protection

- Data were aggregated, kept strictly confidential, and used only for statistical purposes.
- No individually identifiable information included.
- Privacy preserving techniques were rereviewed based on legal frameworks.
- Data were processed on the PURA’s premise.

Internal Migration Use Case

Scan to download the paper ☺

Arai, A., Touray, S., Dibba, L.L., Witayangkurn, A., Shibasaki, R. and Loum, T., **Understanding migration landscape using mobile phone data in The Gambia.** CSAE Conference 2023, 19-21 March 2023. Oxford.

Context

In The Gambia, international and domestic migration provides access to services and opportunities.

1/3 of HHs have at least one international migrant (WB 2019)

55% of migrants send remittances (WB 2019)

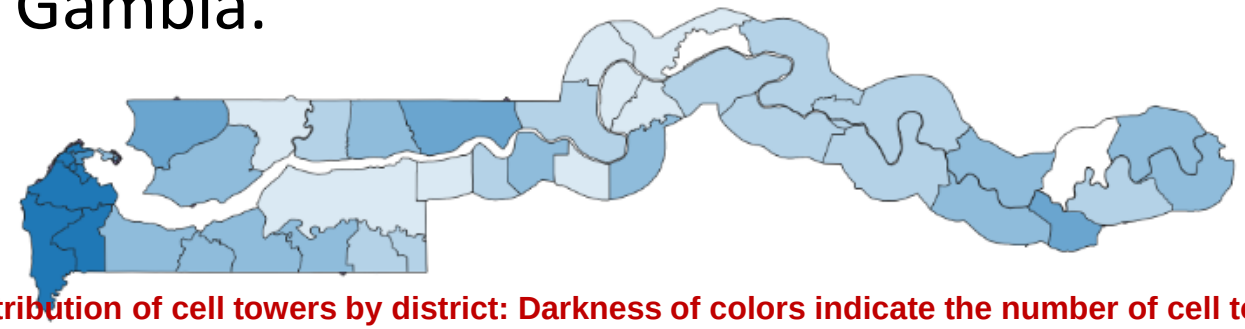
10 Records per day per person on average (2.6 records for calling)

Challenge

- It remains unclear how migration evolves over time
- Better economic opportunities and access to services are pull factors for internal migration … but lack of timely and spatially aggregated data complicates policy response

Data

- We use 16-month CDR data (March 2020 – June 2021) from one of leading MNOs in The Gambia.
- In our dataset we observe:
  - 3.3 million unique IDs in total during the 16 months;
  - 1.1 million IDs, approximately 34%, appear only in a single month.
- Nearly all districts are covered by cell towers we therefore consider that the CDR data cover most of typical subscribers both in urban and rural areas in The Gambia.



Distribution of cell towers by district: Darkness of colors indicate the number of cell towers

Methodology

- We examine the migration landscape in more frequent and geographically granular manner by employing Blumenstock *et al.*(2012, 2022) methods.
- Migration events are defined using changes in the individual's area of residence using the following equation to compute Center of Gravity (CoG):

$$COG_i = \left(1 / N_i \right)_{it}$$

- where an arbitrary point in space within The Gambia is denoted by the vector *r*.
- If an individual *i* makes N_i calls from locations (r_{i1}, . . . , r_{iNi}), that individual's center of gravity is the vector.
- The CoG is returned as the geographic coordinates of cell towers, which are converted to residential location at LGA, district, ward, and cell level.

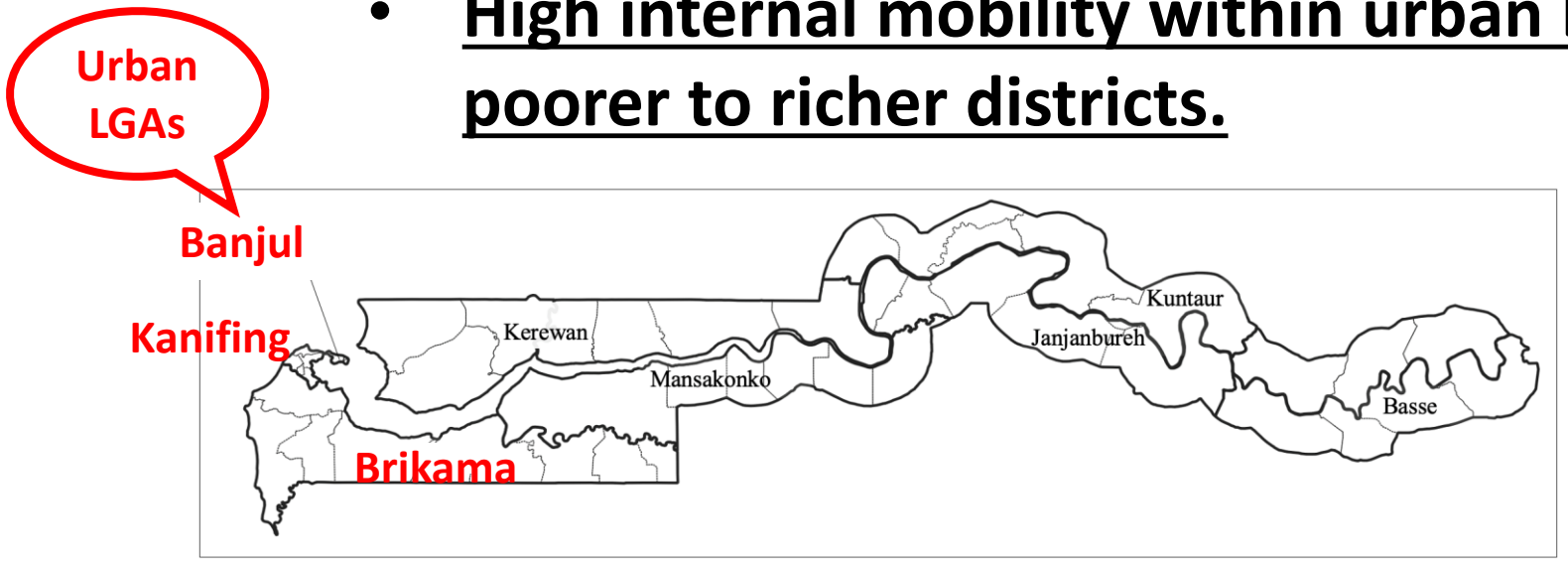
Methodology

- A migration occurs at month *m* if the individual remained in one location for a fixed number of β months prior to m, and was also stationary for β months after and including m but that the locations pre- and post-m were different.
- For the purpose of this study, we set parameters, $m = 2$, $\beta = 2$, to assess if people are stationary before and after migration.

- Based on this definition, we can only use IDs that have records for at least 4 months during the study period:
  - i.e Approximately 40% of the data.
  - Among them, roughly 41,000 IDs (3%) of the whole IDs, are detected as IDs with migration events.
  - 96% of these experienced migration events only once.

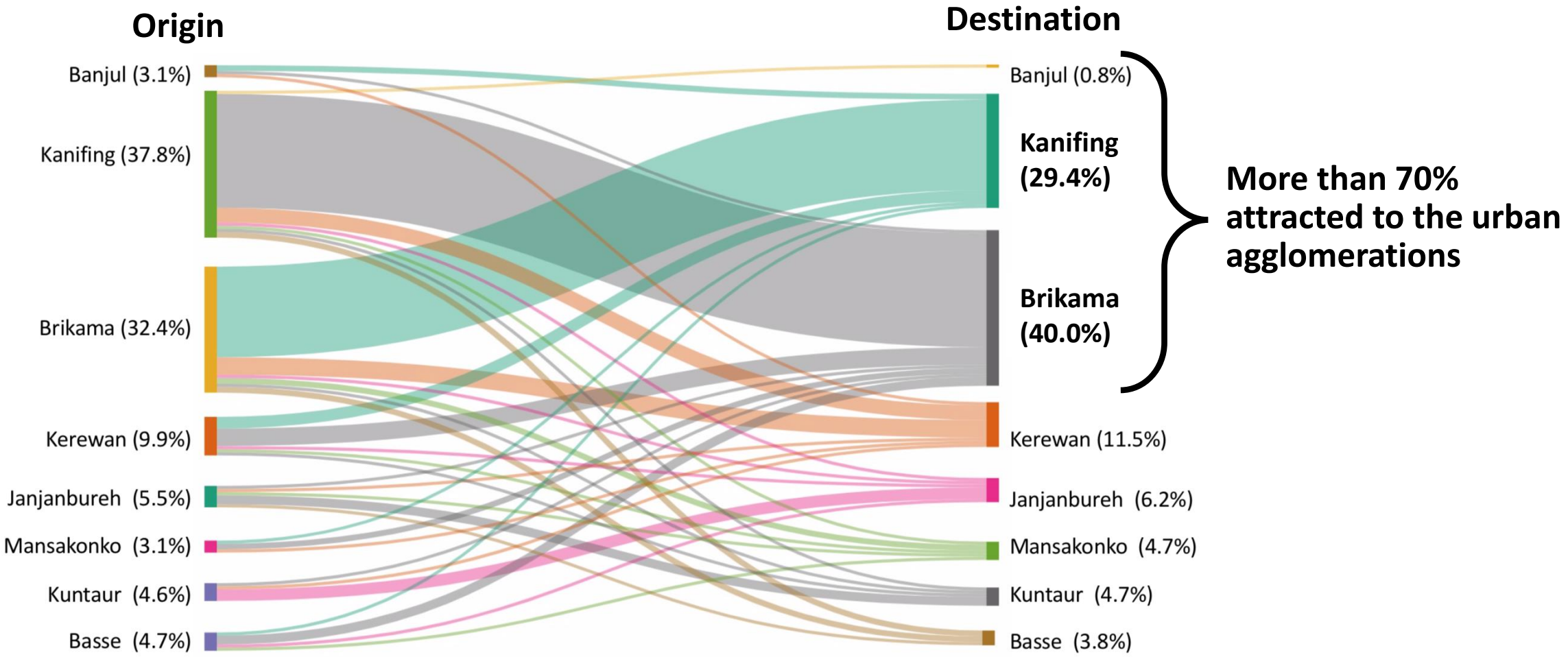
Key findings

- Most migration events occur between **two predominantly urban LGAs** (Kanifing and Brikama)
- We also observe **large number of migration events within LGAs- including in rural LGAs.**
- **High internal mobility within urban LGAs occur from poorer to richer districts.**



8 Local Government Areas (LGAs) consisting of 48 districts

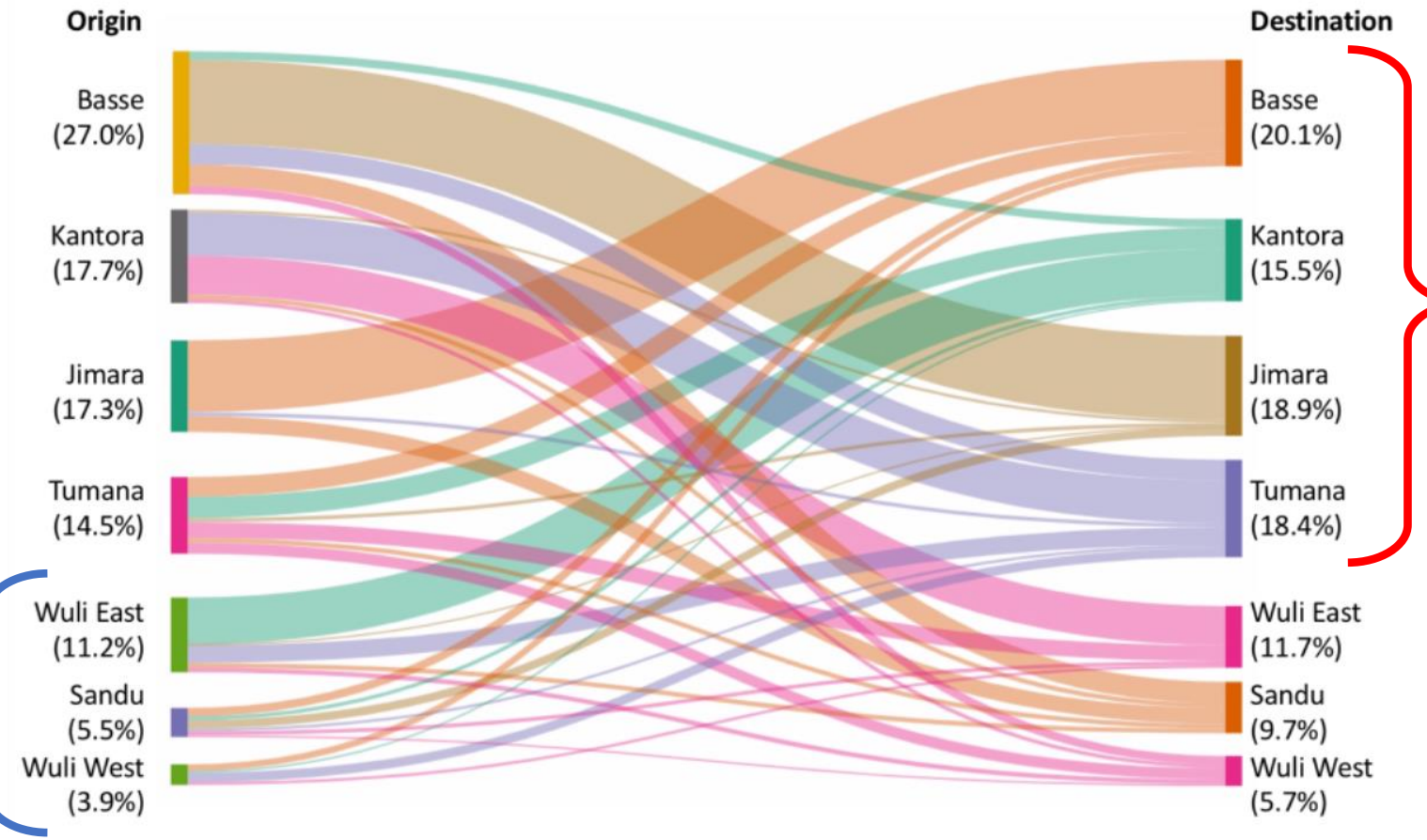
Stylized fact 1:
Services and opportunities in urban agglomerations are probably the main factors attracting migrants.



Inter-LGA migration matrix in The Gambia

Intra-LGA migrations are NOT included. Presented as the percentage to the grand total.

Stylized fact 2:
Migration within rural LGAs is likely motivated by better access to basic services at LGA headquarters relative to other districts.



Small villages in the rural LGA

Large towns in the rural LGA

Migration matrix in Basse LGA (rural LGAs).

Presented as the percentage to the grand total.

Utilizing census data to identify and adjust biases

- Representative biases in CDR data in certain population groups, such as child, elderly, and the poor, woman, lead to skewed analysis results
- Utilizing census data to identify and adjust for bias ensures that CDR analysis accurately reflects the general population.
- Census data provides detailed information on demographic characteristics, including age, gender, income, and geographic location of mobile phone users and non users.

Leveraging census data for adjusting the bias

- By aligning CDR data with census demographics with the breakdown of the mobile phone users and non-users,
- Enhanced accuracy enables policymakers, researchers, and businesses to:
  - access more accurate data for better decision-making
  - improved validity of findings and recommendations derived from CDR analysis

Key takeaways

1. CDR data provide useful insights for understanding internal migration in The Gambia
   - Services and opportunities in urban agglomerations are probably the main factors attracting migrants
   - Economic development and decentralized services may play an important role in increasing economic opportunities and services in rural areas
2. Leveraging census data is expected to improve the representativeness of statistics produced from CDR data