Zhaoyin Chen

Ka Wai Liang

MIST-134

Final Report

Exploring the use of mobile phone data for national migration statistics

**1.Summary**

In this modern world, it is easy and common for people to be moving from one place to another. Migration plays a significant role in determining changes in population size, distribution, and structure. This is the first research that compare individual placed of usual residence across different years to estimate migration flows by matching the definition of migration used in censuses by using call detail records (CDRs). This research paper discusses the use of mobile phone data for national migration statistics. In general, migration statistics are obtained from infrequent censuses or survey and are sometimes inaccurate. The statistics on internal migration are critical for urban planning, infrastructure development, and impact assessment, among other applications; therefore these numbers need to be accurate and up-to-date. To advance the statistics, the researchers attempt to analyze the changes in the spatiotemporal distribution of the mobile phone users through anonymized call detail records (CDRs) to measure the migration at various temporal and spatial scales.

The data is of the regions in Namibia from October 2010 to April 2014, based on a dataset of 72 billion anonymized CDRs. Researchers go in-depth to talk about how they explore and estimate internal migration with precision and accuracy using the data derived and modeled from the CDRs at the subnational and annual scale. This method allows researchers to get better and more accurate results than those from census-derived migration statistics. To make a more accurate estimation of migration, researchers also collated potential migration-related information such as demographic, socioeconomic, geographic, and environmental variables of Namibia. They created three types of model which are CDR-based linear models (CDRLM), gravity-type spatial interaction models (GTSIMs) and lastly CGT SIMs which extended using CDR data form GT SIMs to explore whether CDR data can match traditional migration data.

CDR-based linear models simply using CDR-derived migrating user data alone or combined with covariates used in gravity models. They built four sub-models of CDRLMs using independent variables of CDR-derived migrating user numbers or integrating with other covariates. GTSIMs uses the gravity models which the flow of migration between regions is proportional to their total populations and inversely proportional to the distance between them. The number of regions in Namibia is small (13 regions), they only tested models by replacing the total population variables with the percentage of the population living in urban areas and the precipitation in origin and destination, respectively. To compare the models, they used root-mean-square error, R-squared and Akaike Information Criterion. The best model is the one with the lowest RMSE.

They also talk about how they use CDR’s to assess how migration patterns change over time, using a finer temporal resolution compared with census; how gravity-type spatial interaction models built using CDRs can properly capture migration flows. The researchers are able to obtain promising avenue from the estimation of migration flows using the mobile phone data in an accurate and timely manner.

During the analyze, they find a irregular region on migration pattern which was Zambezi. They were experiencing floods during the year of 2010-2011 which a lot of people moved to other cities and stayed there. So they were excluded from the prediction in the paper.

The study has found that not only can the estimation of migration produced through CDRs be as accurate as census data-derived measures, but also beneficial in terms of updating intercensal migration numbers and understanding the changing in patterns of this annual internal migration. The methodologies presented are very easily implemented when considering the impact of heterogeneous phone ownership across regions and years, and simple linear model built using CDRs results in estimates with high precision and accuracy, with known boundaries to their accuracy in the absence of censuses or surveys.

The basis exists for CDR derived migration data since it cannot count people who do not use smartphones or share smartphones. But the research used data for over years which can eliminate these factors by combining other migration-related covariates and have a significant improvement in the precision and accuracy of outputs from gravity models.

The results of this study have shown that in 2011 Namibia, 2% of the population (40,867 individuals) Namibian migrated by changing their places of residence between region in Namibia over the one-year. The highest migration into was Khomas, the capital region of Namibia, then the highest migration out was Zambezi region of the northeast of Namibia.

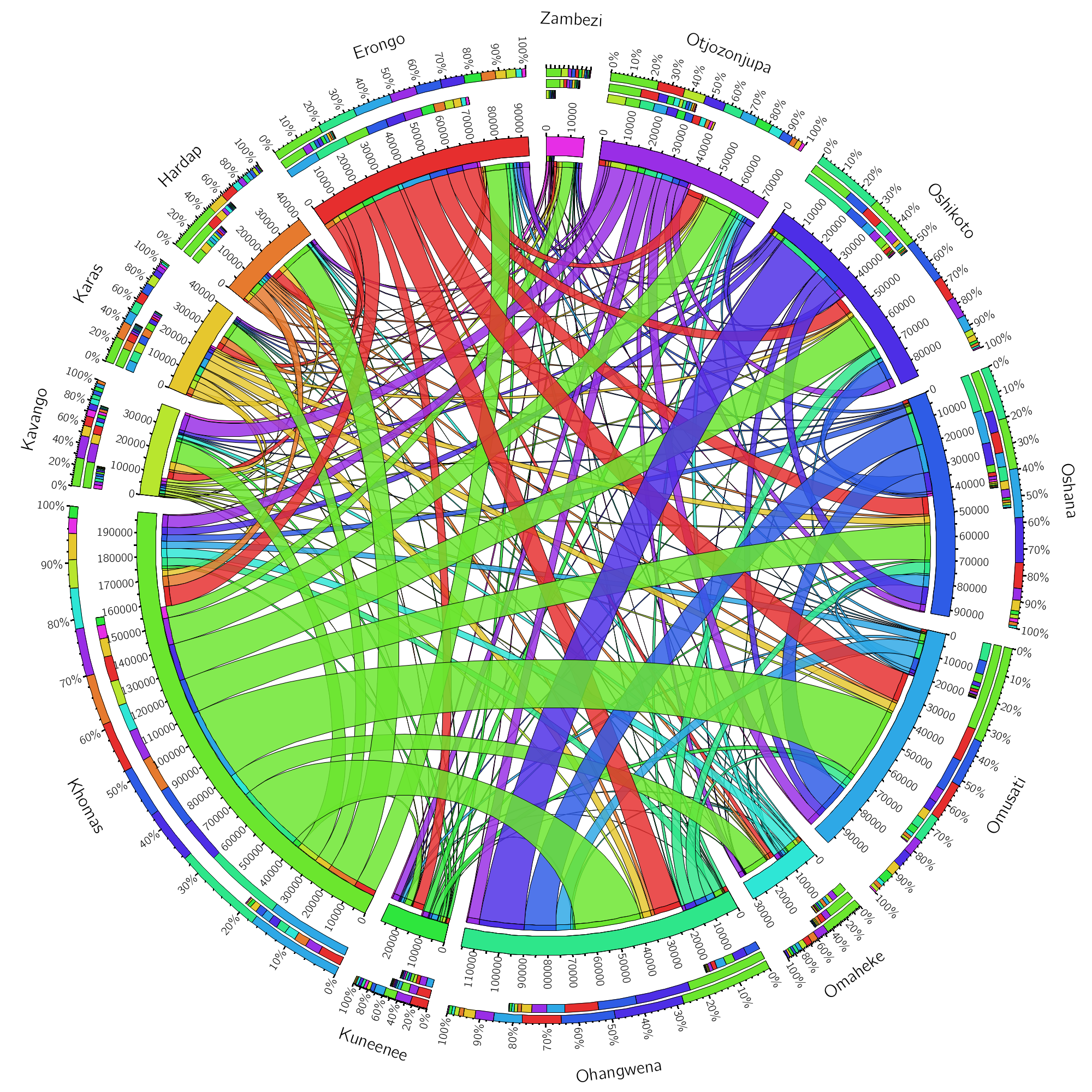
**2. Methods:**

The data comes from the Migration Report from a Namibia Statistics Agency in 2015. The data that we use to create the graphs are: Lifetime Migration from Region to Region, and Short-term Migration from Region to Region – Inflows, Outflows, and Net Migration. Since the data is obtained from a report in a PDF file, we manually enter the one year worth of dataset (.cvs) and import circos to create the data visualization.

As for statistical biases, firstly, the impacts from any seasonal temporary movement should be minimized, such as in December for Christmas. In this study, they use a 12-month time frame to define the migration of residence to prevent bias from temporary movement during seasonal travel. The model also does not take account into a larger number of demographic and geographic group, it only takes place in Namibia.

Most importantly, mobile users do not represent the entire population. Therefore, the CDRs may only provide a partial picture of the migration, excluding those do not have a phone, or are not in the network coverage, or in alternative networks. To be more specific, mobile phone owners typically consist of more educated, and urban males. Mobile network coverage may be substantially lower in remote rural location. Even though the ownership for mobile coverage and usage is increasing, in the study it is shown that a high proportion of SIM card owners and a high share of ownership at the household level.

**3. Data Visualization:** **Reproduce 1** data-driven image from the paper using their provided data. If differences arise in your end result, then discuss the possible explanations for the discrepancy. Provide code notebooks.

  
Figure 1. Lifetime Migration from Region to Region in Namibia in 2011

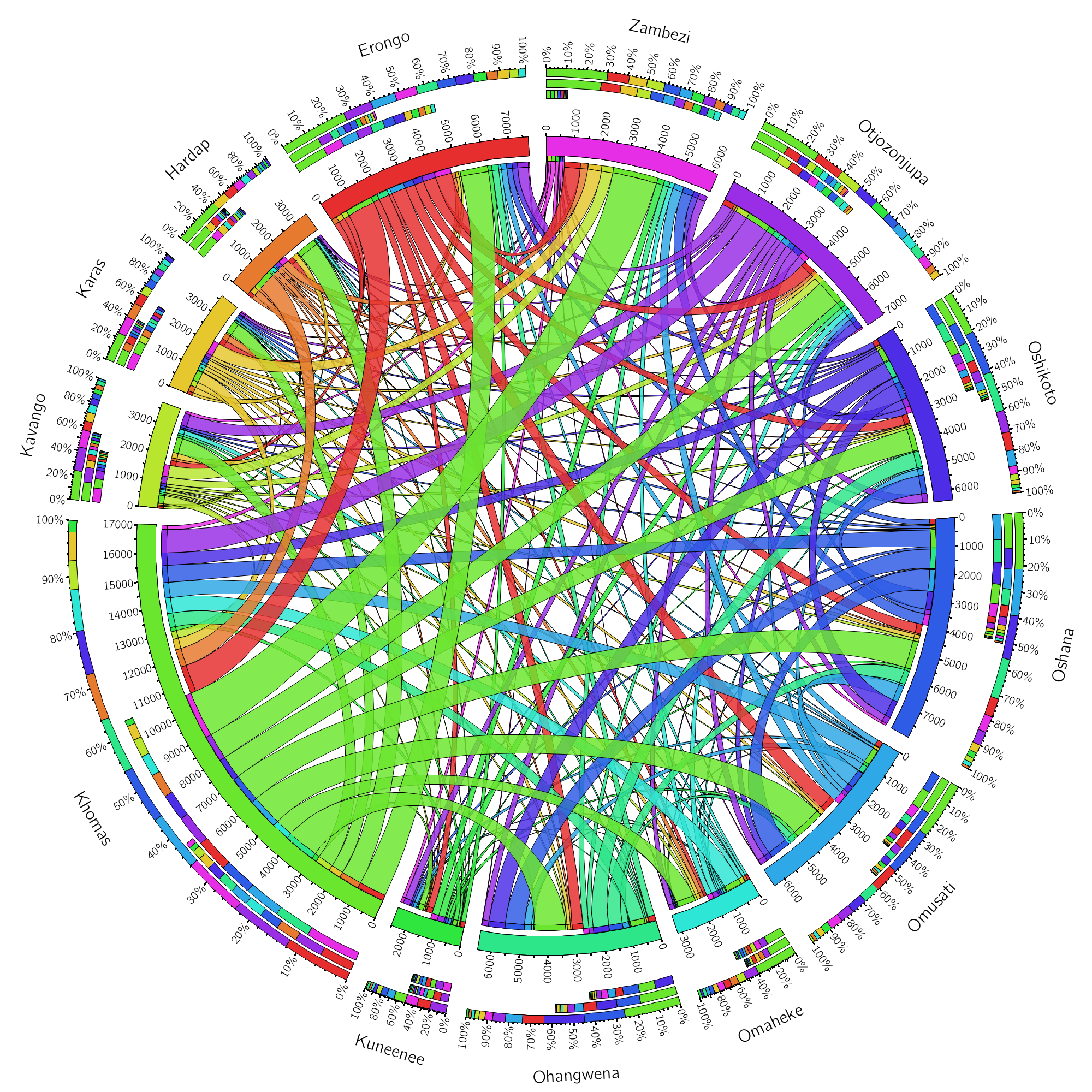


Figure 2. Short-term Migration from Region to Region in Namibia in 2011

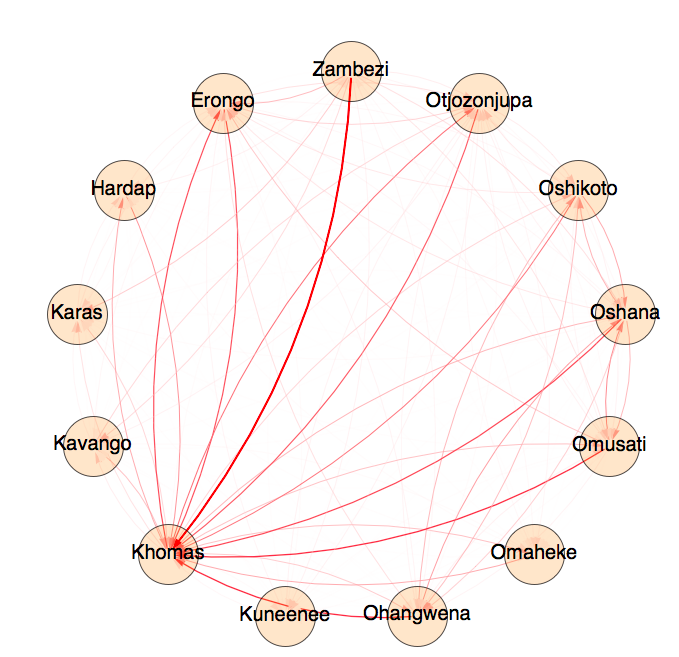


Figure 3. Short-term Migration from Region to Region in Namibia in 2011

We used the data to recreate the image from the paper successfully. The first long term diagram contains the same information where represents how much people migration from one region to another by looking at where the original colors flow. In the original data, there is actually information about migration within the region. Since the paper is focusing on migration among regions, we discard the data. The difference from ours to authors is they you use gap and arrow shape to represent the direction. We just use colors. The graph shows that Khoma is the city that everyone wants to go to.

To improve the graph, we also use short term migration data for the same 13 regions to visualize the pattern. The author did not use this in the report. The short term migration base on the report also matched the long term migration pattern except for a thinner color strip width. Help us identify the