University of California
Santa Barbara

Context, activity, and movement in rural Malawi: using geospatial analyses to abstract spatial insights in the Global South

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Geography

by

Vania Wang

Committee in charge:

Professor Susan Cassels, Chair
Professor Stuart Sweeney
Professor Somayeh Dodge

December 2023

The Dissertation of Vania Wang is approved.

Professor Stuart Sweeney

Professor Somayeh Dodge

Professor Susan Cassels, Committee Chair

November 2023

Context, activity, and movement in rural Malawi: using geospatial analyses to abstract spatial insights in the Global South

by

Vania Wang

To participants from both villages: thank you for generously sharing your time and knowledge. *Zikomo kwanbiri.*

Acknowledgements

A doctoral journey often spans several years of transformative and challenging life events in a student's life. In the space of six years, I've experienced numerous highs and lows. Those who have generously provided advising, financial assistance, and emotional support will always have my deepest and sincerest gratitude.

First, to my dissertation committee members: I would like to thank these mentors for getting me past the finish line of this degree. To Dr. Stuart Sweeney: thank you for your wisdom and insight. Your revisions and commentary consistently reflected your staunch curiosity in my dissertation subject. During times of writers' block and flagging motivation, this was incredibly encouraging. To Dr. Somayeh Dodge: thank you for stepping in as a committee member during the final year of this dissertation. Your subject matter expertise was essential in helping me complete the final chapter of this thesis. And to my ineffably wonderful advisor, Dr. Susan Cassels, thank you. You've been the provider of countless pep talks, immeasurable support, and patience. Thank you for being a mentor, counselor, confidant, and friend.

Further, I am deeply grateful to my parents for leading by example, by instilling in me the importance of critical thought and knowledge. This degree wouldn't be possible without your years of hard work fostering in me my love for exploration and learning. Thank you for courageously immigrating to the United States, building a life in a country that is so foreign to your own, and establishing the foundation of my success and well-being.

Finally, to the love of my life, Angie. I am so confident in your love for me! Why? Because only one person could tolerate reading draft after draft of my thesis, while weathering the many storms of my mood swings and mental breakdowns. Thank you, and I look forward to returning the favor when your PhD comes to a close.

Vita of Vania Wang

Education

expected 2023 — **University of California Santa Barbara**, Santa Barbara, CA

- Ph.D. in Geography
- Thesis: *Context, activity, and movement in rural Malawi: using geospatial analyses to abstract spatial insights in the Global South*
- Advisor: Susan Cassels (Geography)

2015 — **University of California Berkeley, School of Public Health**, Berkeley, CA

- M.P.H. in Infectious Diseases and Vaccinology
- Thesis: *Quantifying catalase levels in patients with active pulmonary tuberculosis and type-2 diabetes mellitus in Mysore, Karnataka State, India*
- Advisor: Lee Riley (Epidemiology and Infectious Diseases)

2010 — **University of Washington**, Seattle, WA

- B.S. in Microbiology with College Honors
- Thesis: *Induction of tertiary lymphoid organs by lymphotoxin alpha and beta*
- Advisor: Gregory Holt (Oncology)

Fellowships & Grants

2019—2021 — Fogarty International Center Health Fellowship
National Institutes of Health, UC Global Health Institute Consortium
$15,000 research support, tuition, stipend, insurance, and travel
Award: NIH/FIC5D43TW009343

2019 — Competitive Innovation Incentive Funds
National Science Foundation
$14,470 research support
Award: DGE1258507

2017—2019 — Integrative Graduate Education & Research Traineeship
National Science Foundation
tuition, stipend, and insurance
Award: DGE1258507

2015—2017 — Allan D. Rosenfield Global Health Fellowship
US Centers for Disease Control & Association of Schools and Programs of Public Health

stipend, training, travel support
Award: 5U36OE000002

Peer-Reviewed Publications

2023 Reid, S. C., **Wang, Vania**, Assaf, R. D., Kaloper, S., Murray, A. T., Shoptaw, S., Gorbach, P., and Cassels, S. (2023). Novel Location-Based Survey Using Cognitive Interviews to Assess Geographic Networks and Hotspots of Sex and Drug Use: Implementation and Validation Study. *JMIR Formative Research*, 7:e45188. Publisher: JMIR Publications Toronto, Canada

2022 Shan, D., Ning, Z., Yu, M., Zheng, H., Yang, J., Gong, H., Li, J., Liu, H., Liu, L., **Wang, Vania**, Ran, X., Han, M., and Zhang, D. (2022). HIV incidence and risk factors among transgender women and cisgender men who have sex with men in two cities of China: a prospective cohort study. *Infectious Diseases of Poverty*, 11(1):26

2019 Shan, D., Su, Y.-y., Yu, M., Yang, J., Li, J., Gong, H., Liu, H., **Wang, Vania**, Han, M.-j., and Zhang, D.-P. (2019). A Comparison of HIV Incidence and Risk Factors Among Subgroups of Men Who Have Sex with Men in Tianjin, China: A Prospective Cohort Study. *China: A Prospective Cohort Study (11/14/2019)*

2019 Wang, L., Podson, D., Chen, Z., Lu, H., **Wang, Vania**, Shepard, C., Williams, J. K., and Mi, G. (2019). Using Social Media To Increase HIV Testing Among Men Who Have Sex with Men — Beijing, China, 2013–2017. *Morbidity and Mortality Weekly Report*, 68(21):478–482

2017 Hu, R., Zhang, F., **Wang, Vania**, Dou, Z., Shepard, C., Zhao, D., Liu, X., Mu, W., Wu, Y., Zhao, H., and Li, H. (2017). Comparing outcomes of HIV-infected Chinese adults on antiretroviral therapy by CD4 count at treatment initiation: A nationwide retrospective observational cohort study, 2012-2014. *AIDS patient care and STDs*, 31(10):413–420

2009 Kalet, I. J., Mejino, J. L. V., **Wang, Vania**, Whipple, M., and Brinkley, J. F. (2009). Content-specific auditing of a large scale anatomy ontology. *Journal of Biomedical Informatics*, 42(3):540–549

Manuscripts In Progress

2019 **Wang, Vania**, Ma, J., Wang, L., Manhong, J., Mi, G., Fei, Y., Shepard, C., Li, Y., Wang, J., Chen, Z., Li, D., Zhang, D., and Hladik, W. (2019b). Estimating the population size of men who have sex with men using a three-source capture-recapture methodology in Kunming Prefecture, Yunnan Province, China. (unpublished)

Scientific Presentations

2020 **Wang, Vania** (2020). *Enumerating the men who have sex with men population in Kunming Prefecture using a three-stage capture recapture sampling activity.* CDC Key Population Surveillance and Estimates Technical Consultation. (oral presentation)

2019 **Wang, Vania** (2019). *WeMap: A methodology to map unplanned settlements using volunteered geographic information.* 7$^{\text{th}}$ Annual Disease Modeling Symposium. (poster presentation)

2019 **Wang, Vania**, Griffith, B., Foko, C., Kawuma, A., and Pearson, C. (2019a). *Using stochastic simulation to inform the design of emergency intervention trials in low- and middle-income countries.* 10$^{\text{th}}$ Annual Consortium of Universities for Global Health. (oral presentation)

2018 **Wang, Vania**, Griffith, B., Foko, C., Kawuma, A., and Pearson, C. (2018). *Using stochastic simulation to inform the design of emergency intervention trials in low- and middle-income countries.* UC Santa Barbara Graduate Simulation Seminar Series. (oral presentation)

2018 **Wang, Vania** and Cassels, S. (2018a). *Formation of sexual ties: determinants of sexual bridging among men who have sex with men and women in Seattle, Washington State.* XXXVIII International Sunbelt Social Network Conference. (poster presentation)

2018 **Wang, Vania** and Cassels, S. (2018b). *Formation of sexual ties: determinants of sexual bridging among men who have sex with men and women in Seattle, Washington State.* Clinic on Meaningful Modeling of Epidemiological Data. (poster presentation)

2007 **Wang, Vania** and Samudrala, R. (2007). *Determining three-dimensional structures using crosslinking and mass spectrometry.* University of Washington 10$^{\text{th}}$ Annual Undergraduate Research Symposium. (poster presentation)

Research Appointments

2014—2015	Graduate Researcher *Quantifying catalase levels in patients with active pulmonary tuberculosis and type-2 diabetes mellitus in Mysore, Karnataka State, India* PI: Vania Wang UC Berkeley, School of Public Health & J.S.S. Medical College
2008—2010	Research Assistant *Immune modulation as a therapeutic strategy for non-small-cell lung cancer* PIs: Drs. Gregory Holt and Mary Disis University of Washington, Department of Oncology
2007—2008	Research Assistant *Clinical Target Volume Project* PI: Dr. Ira Kalet University of Washington, Department of Bioinformatics
2006—2007	Research Assistant *Determining three-dimensional structures using crosslinking and mass spectrometry* PI: Dr. Ram Samudrala University of Washington, Department of Microbiology
2008	Undergraduate Researcher *Water auality assessment of* naulas *in Nainital District, Uttarakhand, India* PI: Vania Wang University of Washington, Department of Microbiology

Public Health Leadership

2013—2015	Board Member Center for Health Leadership, UC Berkeley School of Public Health
2014	Intern Prerana Women's Health Initiative, Mysore, India
2010—2012	Community Health Facilitator US Peace Corps, Efate Island, Vanuatu
2008—2010	Secretary University of Washington, Global Public Health Brigades
2008	Primary School Volunteer Central Himilayan Rural Action Group, Nainital, India

Honors & Awards

2021	Jack and Laura Dangermond Scholarship ($5000)

	UC Santa Barbara, Department of Geography
2019	Ernest Brooks Foundation Scholarship ($250) UC Santa Barbara, Ernest Brooks Foundation
2019	Jack and Laura Dangermond Travel Fund ($600) UC Santa Barbara, Department of Geography
2018	Jack and Laura Dangermond Travel Fund ($500) UC Santa Barbara, Department of Geography
2018	Graduate Student Research and Training Grants in Demography ($175.42, *declined*) UC Santa Barbara, Broom Center for Demography
2017	Unsung Hero Team Award US Centers for Disease Control
2014	Maharaj Kaul Memorial Research Travel Grant ($1,000) UC Berkeley, Institute for South Asian Studies
2014	Lois A. Rifkin Scholarship ($1,500) UC Berkeley, School of Public Health
2014	Center for Global Public Health Practice Fellowship ($4,500) UC Berkeley, Center for Global Public Health
2013—2014	Foreign Language and Area Studies Fellowship (stipend, tuition, insurance) UC Berkeley & US Department of Education
2013—2015	School of Public Health Full Fellowship Award (stipend, tuition, insurance) UC Berkeley, School of Public Health
2010	Microluminescence Teaching Award ($500) University of Washington, Department of Microbiology
2010	Bonderman Travel Fellowship (*finalist*) University of Washington Honors Program
2008	Americorps Service Award ($3,500) Americorps
2005	University of Washington Academy for Young Scholars entrant University of Washington, Robinson Center for Young Scholars

Undergraduate Honors Research Mentorship

2020—2023	Cathleen Li, Department of Statistics and Applied Probability
2022—2023	Sofia Kalopar, Department of Geography

Teaching

2021 | Geography of California (*Geog 148*)
Teaching Assistant
UC Santa Barbara, Department of Geography

2020 | Geography of California (*Geog 148*)
Teaching Assistant
UC Santa Barbara, Department of Geography

2020 | Maps and Spatial Reasoning (*Geog W12*)
Teaching Assistant
UC Santa Barbara, Department of Geography

2019 | UCSB Library Software Carpentry Workshop: Python Thursdays
Instructor
UC Santa Barbara, Software Carpentry

2019 | Introduction to Social Network Modeling in R
Workshop Developer and Instructor
UC Santa Barbara, Broom Center for Demography

2019 | UCSB Library Software Carpentry Workshop: Python, Git, and Bash
Helper
UC Santa Barbara, Software Carpentry

2018 | Physical Geography: Oceans & Atmospheres (*Geog 3A*)
Teaching Assistant
UC Santa Barbara, Department of Geography

2015 | Principles of Infectious Diseases (*Public Health 260B*)
Teaching Assistant
UC Berkeley, School of Public Health

2010 | General Microbiology Laboratory (*Microm 302*)
Laboratory Instructor
University of Washington, Department of Microbiology

2009 | General Microbiology Laboratory (*Microm 302*)
Laboratory Instructor
University of Washington, Department of Microbiology

Academic Service

2018—2019 | Member, Urban Inequalities & Health Disparities Faculty Search Committee
UC Santa Barbara, Department of Geography

2018—2019 | Member, Geography Chair's Graduate Advisory Committee
UC Santa Barbara, Department of Geography

2018—2019 | Graduate Representative, Graduate Student Association

	UC Santa Barbara, Department of Geography
2018—2019	Board Member, Graduate Simulation Seminar Series UC Santa Barbara, Center for Scientific Computing

Media Coverage

2019	*Unplanned Spaces: How movement patterns in rural Malawi impact HIV treatment and prevention*, Access online here UC Global Health Institute Monthly Newsletter
2021	*20/21 Voile Womxn's Backcountry Scholarship Grand Prize Recipient*, Access online here The Voile V-Team Blog

Technical Skills

Computing	R, Python, Java, Javascript, LaTeX
Languages	Mandarin Chinese (fluent), Bislama (fluent), Taiwanese (intermediate), Hindi (basic), Spanish (basic)
Laboratory	bacterial and tissue culture, ELISA, DNA/RNA extraction, PCR, Western Blot, recombinant DNA technology

Professional Affiliations

2019	Certified Software Carpentry Instructor, Software Carpentry

Abstract

Context, activity, and movement in rural Malawi: using geospatial analyses to abstract spatial insights in the Global South

by

Vania Wang

Geography is a crucial dimension of the social determinants of health. Higher resolution data on individual movement and daily activity can offer insights into social and structural disease exposures and various health outcomes. In this dissertation, I ask, what are the individual and community-level activity and movement characteristics of rural Malawians? Further, how can these patterns be characterized in ways that could inform targeted HIV care and treatment? To do this, I implement a two-phased geospatial data collection and analysis protocol in rural Malawi. This protocol was designed to examine the geographical component of the social determinants of health framework within a Global South setting by 1) describing the context in which the activity and/or movement happens, and 2) examining patterns of individual activity and movement. With HIV/AIDS being the leading cause of death in Malawi, this research aims to help reduce geographic barriers of HIV treatment and testing provision. The research protocol described in this thesis may provide a structured methodology to help health service providers visualize, analyze, and understand patterns of mobility and activity within rural villages in the Global South.

I present the application of my research protocol within two rural villages in central Malawi. In the first phase, I conducted gender-segregated guided focus groups from each village. Next, I implemented 199 individual geospatial surveys with village residents. I use latent profile analysis and dynamic time warping to categorize and cluster activity and movement data collected from individual geospatial surveys. The results from these analyses indicate that gender has a magnified impact on individual activity and movement. This research further emphasizes that situational context and community input are crucial components to consider when deploying community-based geospatial research.

Contents

Chapter 1

Introduction

1.1 Dissertation overview and research questions

In rural communities across least developed countries, geographic access is a major barrier to health service delivery and utilization [1–3]. This barrier may include long distances to health facilities from rural villages, unreliable transportation infrastructure, and unmaintained road networks. Geographic barriers can be addressed by understanding where people go on a daily basis, how they navigate their environment, and why they travel. Having a clear understanding of a community's collective activity and movement behavior can help policy-makers understand and eliminate geographic barriers to health access for that community.

In 2015, the United Nations Sustainable Development Goals announced a plan to "provide access to safe, affordable, accessible and sustainable transport systems for all" by 2030 [4]. To assist in this effort, travel behavior research (using travel surveys, journals, or diaries) for least developed countries has served as a means of understanding travel inequities within communities [5]. This includes understanding the social, environmental, and situational contexts that prevent equitable transport access and usage between rural and urban communities. A core inequity lies within the urban and rural divide in least developed nations, where rural inhabitants—unlike their counterparts in urban areas—often lack access to employment opportunities, health services, schools, and/or economic locales. Even though 66 percent of the population in low income countries live in rural areas, there is currently a poor understanding of the transportation and travel barriers for rural communities, and how these barriers affect access to health services [6, 7]. The information collected from travel surveys or diaries could help health providers and policy makers make geographically informed decisions for delivering healthcare to underserved rural communities.

In this dissertation, I consider a unique application of the travel behavior survey in the implementation of HIV prevention and treatment programs. To do this, the research protocol used in this thesis had three main objectives: 1) to collect community and individual-level geographical data from residents in rural settlements, 2) to analyze this data for meaningful patterns in villager activity and movement, and 3) to suggest geographical insights relevant for policy makers working within the domain of HIV implementation science in resource poor settings. In this thesis, I explore analytical techniques that interrogate how individual geographies—specifically daily activity and movement—can impact personal access to health services. My research protocol also adds two elements that are commonly missing from the traditional travel survey. First, I incorporate a qualitative focus group phase prior to implementing individual travel surveys, which adds a community-driven component to help inform and develop survey and research protocols. As a U.S.-based researcher working in a Global South context, these focus groups were critical in facilitating an exchange of information that helped guide and plan the research. Secondly, I add a visual and graphical component to these survey activities, by asking study participants to list their daily travel in a tabular format, and draw their travel trajectories on a printed satellite image of their village. Both of these elements were combined to layer additional cultural, social, and geographical context onto the data collected from the travel survey.

This dissertation aims to explore overall activity and movement behavior patterns in rural Malawi that may be relevant for strategizing targeted HIV treatment and prevention efforts. In doing so, I answer three groups of questions about human activity and movement within the scope of HIV implementation science in Malawi. The first set of questions, addressed in Chapter 4, concerns village-level patterns and trends in activity and movement:

1. *What characteristics of village life (e.g. demographic composition, economic activity, community health issues, and seasonal changes) are salient to understanding overall activity and/or movement among village residents? How can these insights be used to frame the results of geospatial analyses applied in rural communities in the Global South?*

The second set of questions, addressed in Chapter 5, concerns individual daily activity and energy expenditure:

2. *Are there distinct archetypes of activity among village residents? If so, what are the key characteristics that define them?*

The third set of questions, answered in Chapter 6, concerns individual daily movement trajectories:

3. *Are there distinct patterns or clusters of movement among village residents? If so, what are the spatial and demographic characteristics that separate and distinguish these clusters?*

1.2 Dissertation structure

The dissertation begins here with an overview of the motivations behind this project, its guiding research questions, and how it may address gaps in the field of health geography. In Chapter 2, I describe relevant background literature that contextualizes this research. I also describe the theoretical and conceptual frameworks that ground this dissertation's methods and findings within the field of geography. Chapter 3 describes the study setting, data collection methodology, and survey instruments used in this research.

Chapters 4, 5, and 6 describe the individual analyses applied to address the three aforementioned research questions. Chapter 4 details the findings from my initial guided focus groups, which establish the groundwork for the analysis in Chapter 5. In Chapter 5, I describe the implementation of latent profile analysis on individual-level activity data. In Chapter 6, I detail how dynamic time warping was used to cluster the individual travel trajectories collected from our participants. Finally, Chapter 7 summarizes the results from the three substantive analytical chapters, and reflects on the overall implications of these findings within the domains of movement and activity research and health geography.

1.3 Addressing gaps in health geography

This dissertation explores the application of two methodological and analytical approaches in health geography. First, this dissertation outlines a step-by-step sequence wherein village-level focus group data describing community movement was first collected, followed by the collection of individual-level activity and movement data. By using the set of analytical tools presented in chapters 5 and 6, this geographical data reveals dominant patterns of daily activity and clusters of movement behavior among the village residents. These insights may be useful in building access to health service provision at these village sites. Further, even though the proposed data collection methodology was specifically applied to two village sites in central Malawi, the methods and analyses presented in this dissertation may be applicable to other communities in the Global South.

Second, this dissertation uses classification and clustering analyses on two sets of geographical data: individual activity and movement trajectory data. To my knowledge at the time of writing, these analytical approaches have not been applied to rural residents in Malawi, or indeed, the Global South. The results of these analyses reveal key insights about addressing the specific needs of rural Malawians within the domain of HIV prevention and treatment.

Chapter 2

Background

2.1 Introduction

In this chapter, I discuss the existing research and literature that guided my research question, methods, and analyses. First, I introduce the core theoretical framework that grounds the motivation behind developing this research: the relationship between the social determinants of health (SDoH) framework and HIV acquisition. Although biological factors do contribute to the transmission of HIV, SDoH underpin the circumstances in which infection occurs [8]. The data collection and analysis protocol developed for this thesis was designed to contextualize the geographical component of the SDoH framework, especially how it pertains to HIV transmission in rural Malawi.

Next, I provide an overview of Malawi, including the state of the country's HIV epidemic and the status of infrastructure in rural areas. Here, I highlight country-specific information that contextualizes the study's analyses and research questions. Specifically, I outline the challenges Malawians face in terms of the HIV epidemic and barriers rural Malawians encounter with respect to infrastructure and transportation.

Finally, I discuss the placement of this research within the field of geography. The analyses presented in this thesis traverses several different sub-fields of human geography, including health geography, travel behavior research, and movement trajectory analysis. Here, I split this section of the background into three subsections that individually provide context for the research questions presented in Chapter 1. First, I describe participatory and community mapping as a means to gather collective knowledge about a community's history, demographics, and geographical environment. Second, I review travel behavior research, time-use data, and human activity tracking, and explore how these form the conceptual framework for understanding the latent profile analysis (LPA) implemented in Chapter 5. Finally, I introduce the field of trajectory analysis, which provides a set of tools for collecting, analyzing, and comparing movement trajectories of animals and people. I also discuss different trajectory similarity measures, and why dynamic time warping was chosen as the clustering similarity measure used in Chapter 6.

2.2 Social determinants of health

The foundational theoretical framework for this research activity is the social determinants of health. Poor access to health resources, economic disparities, and gender inequalities can contribute to and exacerbate infectious diseases like HIV. Such environmental, geographical, and social risk factors contribute to the ongoing marginalization of specific communities while contributing to poor health outcomes among these groups [9–11]. The intersection between diseases and their causes must be approached through the theoretical lens of Social Determinants of Health (SDoH) [12]. The SDoH framework emphasizes contextual (distal) risk factors over immediate (proximal) causes of disease.

Proximal risk factors are immediate or close precursors to disease or ill health. Examples include ongoing and persistent exposure to violence in relation to trauma or injury, environmental hazards like poor access to sources of clean water with respect to cancer risk, and current tobacco use as a precursor to lung and heart disease [12, 13]. With respect to studying the influences of HIV acquisition, some proximal risk factors that increase risk of infection include injection drug use, anal sexual intercourse, and receiving unsafe blood transfusions [14].

Distal determinants of health include the national, institutional, political, and cultural factors that indirectly influence health by acting on proximal factors. That is, distal factors comprise the context in which proximal determinants are constructed and are the most difficult to change, transform, or influence. Examples of distal factors include, social and political factors that contribute to poor access to clean drinking water; social pressures that encourage tobacco use; or impacts of colonialism or racism that have historically prevented impacted communities from accessing potentially beneficial social or health services.

Another important distal factor is geography. Geography plays a critical role in the study of infectious diseases and their epidemiology. Within the fields of public health and health geography, the rural-urban divide is a common dichotomy that is well-studied within the SDoH framework, both within industrialized and Global South contexts [15–17]. For example, residents of rural and geographically isolated communities in the Global South have poorer access to health education [17]. Neither geographic isolation nor access to health education alone cause HIV infection, but in tandem, they create a situation in which an individual's risk of acquiring HIV increases. Compared to their urban counterparts, rural inhabitants often lack access to electricity, consistent and clean water sources [18, 19], and economic and social opportunities [20], all of which are associated with geographic barriers and isolation. Furthermore, rural residents in least developed countries (LDC) are less likely to seek maternal health services [21] and have poor access to other healthcare services [22], including HIV treatment and prevention services. In many LDCs, rural dwellers often outnumber urban populations. Collectively, 65 percent of the total population in all LDCs live in rural communities with fewer than 300 inhabitants per one km2 spatial unit [6, 23].

If successfully addressed and targeted, changes in distal determinants of health may yield the greatest health impacts and, thus, long-term change towards systemic health inequities. Furthermore, given the known differential access to health and infrastructure services between rural and urban locales, geography is a critical distal factor when examining factors that can influence individual and collective health outcomes [13, 24, 25]. Without careful consideration of geography as a distal risk factor for HIV risk, infection prevention strategies are unlikely to succeed. In this dissertation, I dissect the geographical component within the SDoH framework beyond the rural-urban impact on health access, and hone-in on individual geographical heterogeneities in mobility and activity. Before delving into an examination of literature that explores these individual-level geographies, I introduce the place where my study was implemented and conducted: Malawi.

2.3 The HIV epidemic in Malawi

In 2022, there were over one million people living with HIV in Malawi out of a population of 20 million [26, 27]. The prevalence of HIV among adults aged 15 to 49 is estimated at 7.1 percent [27]. Epidemics are said to be concentrated if transmission occurs largely in vulnerable groups and key populations such as men who have sex with men, sex workers, or people who inject drugs. Conversely, epidemics are termed generalized if transmission is sustained by sexual behavior within the general population; and would persist despite effective programmes for key populations. Typically, generalized HIV epidemics are defined on the basis of population prevalence of HIV greater than one percent. In general, most HIV epidemics in Latin America, the Middle East, Europe, and Asia are concentrated, while the epidemics in most of southern and eastern Africa are generalized.

In December 2020, UNAIDS released a revised set of ambitious targets calling for 95 percent of all people living with HIV to know their HIV status, 95 percent of all people with diagnosed HIV infection to receive sustained antiretroviral therapy (ART), and 95 percent of all people receiving ART to have viral suppression by 2025 [28]. Known as the "three 95s", these targets were revised from the "three 90s" goal announced in 2014, with a similar aim in testing people living with HIV (PLHIV), recruiting them into consistent treatment, and aiming for sustained viral suppression. These global targets have been widely adopted and largely successful in placing emphasis on the scale up of ART, especially in countries with generalized HIV epidemics. Malawi is one such example: although Malawi's HIV epidemic is still generalized, significant progress has been made over the past decade in recruiting PLHIV into the cascade of care and treatment. Between 2010 and 2020, the annual number of new HIV cases decreased from 56,000 to 21,000 new infections, AIDs-related deaths decreased from 36,000 to 13,000, and the coverage of PLHIV receiving ART increased from 28 percent to 85 percent [27].

Despite high HIV burden and health system constraints, Malawi built a remarkably successful HIV treatment program by scaling up treatment with a focus on ensuring equitable access to ART across the country [29–32]. In 2004, the Malawi Ministry of Health decided to scale up ART nationwide. Through concerted and strong national leadership, the Ministry of Health in Malawi was able to implement standardized systems for staff training, accrediting health facilities, and delivering and monitoring ART. At the start of the treatment scale up in 2004, only nine hospitals were providing ART across the nation. By the end of 2015, there were 716 ART clinics nationwide. The system for delivering treatment became decentralized from the nine original hospitals, thereby providing testing and treatment services that many rural Malawians were able to access at satellite health centers.

Although national indicators show that Malawi is making remarkable progress in controlling its HIV epidemic, subnational trends show disproportionate and heterogeneous decline in HIV prevalence, highlighting an existing need to consider regional or geographical context within epidemiological research [33]. Furthermore, it is well documented globally and within the sub-Saharan Africa region that men are more likely than women to die of AIDS-related causes, less likely to know their infection status, and less likely to be on ART treatment [34–36]. With this documented heterogeneity in accessing HIV services between demographics and geography, the following questions therefore remain: 1) How can patterns in daily activity and movement help researchers extract inferences on how to reach the lingering PLHIV who do not know their status? 2) How can geographical data and geospatial analytical tools help researchers and policy makers better understand how people access and utilize health services?

2.4 Infrastructure in rural Malawi

In Malawi, 85 percent of the country's population is considered rural [37]. In addition to population density, rural life is often defined by transportation access and reach. However, there are not clear definitions of "rural" vs "urban", and thresholds vary considerably between regions and communities. Given improved access to information and communication technologies, and seasonal or long-term migration to cities or factories, it becomes increasingly difficult to define and clarify differences between rural and urban contexts. Notwithstanding these caveats, most rural dwellers are engaged within the agriculture sector, and many depend on subsistence farming for their livelihood.

When looking at the data in terms of rural-urban disaggregation, rural poverty rates in Malawi are significantly higher compared to its urban areas, where one fifth of the rural population (20 percent) lives in multidimensional poverty compared to 4 percent of the urban population [38]. The multidimensional poverty index refers to deprivations in three dimensions, each having its set of specific indicators: health (nutrition, child mortality), education (years of schooling, school attendance), and living standards (cooking fuel, sanitation, drinking water, electricity, housing, and assets) [38]. In general, disparities and inequalities between urban and rural contexts are on the rise globally [39].

With respect to transportation infrastructure, 26 percent of Malawi's existing roads are paved as of June 2016 [40]. The rural access index for Malawi stood at 23.1 percent in 2016, leaving 11.3 million rural residents disconnected to paved roads that are in fair or better condition [2]. Inter- and intra-village travel is often facilitated by walking or bicycle taxi, while longer distance travel (e.g. traveling for health procedures at the regional health center) can turn into whole-day or multi-day excursions. Travel in rural Malawi is often complicated by the changing nature of physical paths, walkways, and unpaved roads, which can appear and disappear in response to weather conditions, seasonality, and other structural changes in the village. Similarly, most rural Malawian villages are informal rather than pre-planned by urban planners and civil engineers. As shown later in the results from the focus groups organized at the study villages, they develop in accordance to tribal identity and/or familial or communal relationships. In this thesis, I introduce a set of geographical analytical tools that explore individual spatial heterogeneity of residents in rural Malawian villages. These analyses are aimed at further exploring the range of activity and movement of rural Malawians, to elucidate the relationship between individual geography and health provision needs of rural villages.

2.5 The role of geography in studying HIV epidemiology

In this dissertation, I hone in on the SDoH framework by taking a closer look at geography as a distal determinant of health; specifically the variations and patterns of individual activity and movement. Place-based factors alone can greatly influence health outcomes [24, 41, 42]. Within HIV treatment and prevention implementation science, geographical techniques have helped researchers consider spatial disparities, which impacts access to healthcare like HIV testing and treatment. Geographical approaches to HIV treatment and prevention have examined differences in testing coverage between geographical units, which include neighborhoods, US Census-designated geographies, or any other defined spatial unit [43]. In prior studies, HIV key population sizes were compared based on geographic and spatial boundaries [44–46]. These analyses operate on the assumption of static space, or predefined spatial boundaries or units, like counties in the United States or other hierarchical standardized administrative units. These units become associated with certain health-related indicators, such as the number of incident cases of HIV across subnational districts in Tanzania. Missing from these analyses is the consideration of how individual activity and movement behavior could affect variation in HIV cases across districts. For instance, some questions that consider individual-level geographies are: 1) how many cases in a given district in Tanzania are contributed by individuals from external districts? 2) What are the individual activity and movement characteristics of these non-resident HIV cases? And, 3) how could they be targeted and connected to the cascade of HIV care and treatment?

Beyond a static definition of space, a more flexible definition of individual geography is an individual's activity space [47–51]. The concept of activity space has been defined as "the subset of all locations within which an individual has direct contact as a result of his or her day-to-day activities" [52]. Epidemiologists and health geographers have argued that focusing solely on defined space—neighborhoods, counties, districts, etc—may limit the consideration of the full range of influences on a person's health outcomes [51]. Characterizing the space within which people move or travel during the course of their day-to-day activities (their activity space), may provide a more comprehensive and accurate assessment of the heterogeneity of exposures that an individual encounters. In turn, this could help to elucidate the mechanisms by which geographic environments affect a person's health. Although there are numerous studies that link activity space and HIV infection risk/transmission among communities in developed contexts [53–55], there is a dearth of research studying these mechanisms in rural or least developed communities.

While seeking to understand and describe an individual's activity space, their overall spatial context (i.e. cultural, political, and social environment) is a core component for consideration. Spatial context can be complex and difficult to assess, especially as it is embedded within layers of mutually-influencing situational, geographical, social, environmental, and cultural forces. For example, in some research studies, place (i.e., context) and its residents are perceived as independent explanations for health inequity, without consideration for the relational dynamic between these factors [56]. Scholars have called for a reconfiguration of space in health studies, in favor of conceptual models that consider geography as the underlying thread for health-influencing variables [42,57]. Although it's challenging to distill these multi-scale and multi-system complexities into patterns, this thesis aims to focus solely on the geographical component of the SDoH framework. Specifically, I examine dynamic and changing space, by studying the activity and movement patterns among residents in a rural Global South context, with HIV as the health outcome of interest.

2.6 Conceptual framework of HIV, human movement and activity

In this dissertation, I distinguish between three components of individual geography: activity, movement, and spatial context. These components are extended from the aforementioned activity space concept, or the places and spaces where a person occupies over the course of their day [50, 51]. There is some variation in how prior studies defined the structure of an activity space, but most definitions include activity and movement as two major spatial components: activity being the the daily locations visited and exertion expended at those locations; and movement being the path of travel between these daily locations [51, 52]. To this existing definition, I add spatial context, or the socio-cultural environment in which an individual operates, as an additional component to an individual's activity space. In summary, I posit that activity, movement, and spatial context constitute the totality of an individual's activity space. In the next three subsections, I further clarify and define the ways in which each substantive chapter examines spatial context, movement, and/or activity.

2.6.1 Chapter 4 – Participatory and community mapping

In Chapter 4, I use focus groups and participatory mapping (or, community mapping) to collect crowd-sourced geographic and contextual data about this study's village sites and their residents. The main purpose of focus group research is to draw upon the respondents' beliefs, experiences, and collective knowledge in a way that would not be feasible using other methods, like individual surveys or interviews. The interactive aspect of focus groups provides an opportunity for respondents to explore different points of view, and formulate and reconsider their own ideas and understanding. Compared to individual interviews, which aim to obtain individual ideas and experiences, focus groups elicit a multiplicity of views and emotional processes within a group context [58–60]. Within public health research, focus groups are often included in mixed-methods studies to gain more information on refining or constructing questionnaires or interpreting analytical results [61]. Focus groups can be a powerful means for study designers to understand the context behind a study setting or research topic.

Similarly, participatory mapping is a tool for understanding how participants collectively view and define their village community, and could be used as a *de facto* assessment for the community's spatial awareness and understanding of their own community [62–65]. Depending on resources and availability of materials, participatory maps can be created on large pieces of paper or digitally through mapping applications and GPS devices. There are variations towards implementing a participatory mapping activity, but in general, participatory mapping is a map-making process that attempts to visualize the association between land, place, and the local communities that reside there.

As with any type of map, participatory maps can depict detailed information of village layout, landmarks, and/or infrastructure (e.g. water sources, roads, or the location of landmarks or homes). However, participatory maps are not confined to only presenting locations and features; they can also illustrate important social, cultural, and historical knowledge. Depending on the survey themes and research questions, participatory mapping is a flexible and powerful technique to understand the contextual geography of a survey population. Participatory mapping enables villagers to display local knowledge and information about their community, while creating a sense of ownership with the space they inhabit. In this application, the combined focus group and community mapping activities helped generate rapport between community members and the visiting researchers, introduced the multi-stage research activity towards the community, and generated core insights and themes about village life.

2.6.2 Chapter 5 – Observing, measuring, and capturing daily activity

In Chapter 5, I use travel surveys as the data collection tool for my analysis. Travel behavior research investigates fundamental questions about human activity, including why, how, when, and to where individuals move, and how their activity is linked to their contextual environment [66]. Within this framework, time-use research is an interdisciplinary field aimed at understanding an individual's allocation of time during a given period, and the factors that influence time-use choices [67]. Research questions that can be addressed using time-use studies include, "How does an individual's lived environment influence their daily time composition?"; "How do sociodemographic factors influence a person's time-use"; "What can a population's overall time-use tell health providers about implementing HIV prevention protocols?". *Time-use epidemiology* can, therefore, be defined as the study of determinants, distributions, and effects of health-related time-use patterns in populations [68].

Most travel surveys are administered retrospectively, using a diary format that asks the participant to recall trips they took on a chosen day. The diary survey first appeared in the late 1970s, with a booklet design developed by Socialdata in Germany [69]. This survey design was then introduced in the United States in 1982, and used in surveys in various US cities in the early 1980s [70]. The trip diary became the most popular design for household travel surveys in the 80s, and continues to be used at present. Travel surveys have mainly been applied to urban transportation research in developed countries or high-income communities, and similar models are assumed to apply in the Global South. However, these assumptions are met with unique challenges when implementing travel surveys in the Global South. An example is, low income countries often have a more pronounced class divide that influences differential access to transportation services and use of transport modalities [71,72]. Therefore, to identify specific geographic barriers and challenges faced by individual communities, travel surveys need to be readapted prior to being applied towards communities in the Global South, specifically rural locales.

With the decision to conduct time-use analyses with travel behavior surveys comes the key question of *how to* classify daily activities. In existing time-use epidemiological research, by far the most common schema is to classify activities into sleep, sedentary, physical activities [73,74]. Physical activities can be further subdivided into low-physical intensity (LoPA) and moderate-vigorous intensity activities (MVPA). Three characteristics are used to classify an activity into sleep, sedentary, LoPA, and MVPA: 1) wakefulness (awake or not awake), 2) posture (lying, sitting, or standing), and 3) energy expenditure ($<$1.5 metabolic equivalents [MET] and $\geq$ 1.5 MET) [75]. Sleep generally is the lowest on the energy expenditure spectrum, with an average of 0.95 MET [76]. A person engaged in a sedentary activity is usually awake, lying or sitting, and expending low energy ($<$1.5 MET) [77]. LoPA activity is characterized by a person being awake, standing, and expending less energy (1.5 - 3 METs) while a person engaging in MVPA is awake, standing and expending over 3 METs [78–81]. For the purposes of labeling activities in this research, I classify a person quietly standing as LoPA.

The sleep-sedentary-LoPA-MVPA schema is the most widely used classification criteria in health and activity research, and is most applicable in the sphere of studying daily physical activity. Within the fields of epidemiology and public health, there is a dearth of research that explores differences in activity classifications beyond sleep, sedentary behavior, and differing levels of physical activity [73, 74]. In this dissertation, physical activity and exertion can certainly contribute to differences in activity classification, but isn't the sole determinant of classification differences. In addition to the sleep-sedentary-LoPA-MVPA classification, I explore different classification schemes in this study, and determine which is the most appropriate given this research application, population, and cultural context.

In Chapter 5, I use latent profile analysis to classify daily activity behavior of rural Malawians into *activity archetypes*. By *activity archetypes*, I mean latent patterns in the participant's time-use data that show overall patterns of daily activity and behavior within the chosen village sites. Activity archetypes may be applied to categorize these multidimensional activity patterns, which may allow health providers to tailor specific interventions for each activity archetype. For example, if one activity classification is found to engage in leisure activities more than another, I can further ask questions about the 1) the demographic composition of this classification group, 2) the types of leisure this group engages in, 2) where these leisure locations are, and 3) whether or not these locations are feasible contact points for health delivery. When planning health interventions, researchers may therefore apply these geographical insights to inform *how*, *when*, *where*, or *to whom* interventions are applied.

2.6.3 Chapter 6 - Capturing and defining movement trajectories

On a day-to-day basis, most individuals move around to perform their routine activities. We move around our homes, neighborhoods, towns, and rarely stay within a predefined spatial boundary or geographic place. In fact, an individual may spend the majority of their time outside their home or neighborhood. This movement is key to understanding public health service access and disease risk. Not only are individuals likely to seek social or health services outside of their immediate home area [68]; through movement, individuals experience and may come under the influence of a variety of environments and disease exposures [82]. Given that human movement patterns are not static, any examination of how humans interact with their environments must be adaptive and respond to their movement.

Despite the ubiquity and relevance of human movement to public health research, movement data can be challenging to capture and analyze, specifically in resource poor settings. At present, most movement data is represented in the form of *trajectories*, defined as “a sequence of time-stamped locations” [83], or, “as a discrete trace on which a moving object travels in geographical spaces” [84]. Trajectory data can be grouped into *explicit* or *implicit* categories: explicit trajectory data provides time and location information with strong spatiotemporal continuity, while implicit trajectory data has weak spatiotemporal continuity [85, 86]. Explicit trajectory data is usually actively collected using GPS devices, while implicit trajectory data is collected more passively using RFID, Bluetooth, WiFi, or other signal-based sensors [85, 86]. Given the relative expense and cost of acquiring GPS-enabled devices in Malawi, I explore using cost-effective means to collect individual movement and mobility data. In Chapter 6, I describe collecting and analyzing an analog of explicit trajectory data, collected through drawn pathways on paper maps.

After collecting trajectory data, the next logical step is to derive insights from the data using a selection of trajectory data mining methods [86]. In Chapter 6, I describe using a clustering technique to group trajectories into distinct categories. Trajectory clustering is an unsupervised learning process that reveals patterns and insights within a trajectory dataset by sorting trajectories based on *similarity* [86]. Many measures exist for calculating the similarity between two trajectories, each with their own strengths, weaknesses, and applications. The concept of similarity can be described by three intuitions: 1) the more features two objects share, the more similar they are, 2) the more differences there are between two objects, the less similar they are, 3) two objects are maximally similar if they are identical in all respects [87]. Similarity can be quantified using *similarity measures*, represented by mathematical functions used to numerically compare objects. For many similarity measures, the inputs are two objects and the output is a number, usually a value that's used to quantify the cost of transforming one object to another [88]. This cost is usually represented by a distance measure, or, inverse of a distance measure.

The adaptation and usage of trajectory similarity measures have been liberally applied within time-series analysis [89] and geometric shape matching [90]. Within applied science, trajectory similarity measures are used in ecology to understand animal movement and behavior [91, 92]. However, these measures and techniques have not been applied within epidemiological science or public health. For the purposes of this dissertation, I am interested in identifying and explaining movement patterns from a sample of rural Malawians that may be salient within a public health application. Comparing the similarity of two or more instances of human movement trajectories, specifically within the dimensions of geographic space, is one way to achieve this purpose. How, then, are movement trajectories compared and measured?

There are a plethora of trajectory similarity measures available, but some are used more than others, including Dynamic Time Warping (DTW) [93], Euclidean Distance [94] (EuD), Longest Common Subsequence Distance (LCSS) [95], Edit Distance (ED) [96], and Fréchet Distance (FD) [97]. Of these common trajectory similarity measures, DTW was the best choice for this application based on the following reasons. First, the EuD measure could be eliminated from these choices because it is a one-to-one point-matching comparison of trajectories, requiring them to be of equal length. This limits its usage when analyzing human movement trajectories, because human trajectories often present variations that result in different lengths. Similarly, the ED measure is not an appropriate measure. Although the ED measure does not require trajectories to be identical in length, the algorithm imparts a penalty for two trajectories of different lengths, even if one is a subsection of the other.

Second, in my dataset, two trajectories that traverse the same paths through identical geographic regions could have a very different set of coordinates, because this trajectory data was digitized and traced manually without a consistent sampling frame (instead of formed through aggregating points sampled regularly from a GPS device). This eliminates LCSS as a suitable measure, because this algorithm is better suited towards trajectory data with consistent sampling frames. Finally, the FD similarity measure could be a suitable algorithm for this application. The algorithm is capable of handling trajectories of different sampling frames and lengths. However, because it is the most computationally intensive out of all the methods listed (with the exception of the Euclidean Distance measure), I decided on using DTW for the clustering activities applied in this dissertation [98].

In Chapter 6, I will describe how the DTW algorithm works with partitional clustering to group and categorize trajectories, in an effort to resolve overall themes in movement among rural Malawian villagers. The goal of this chapter is to present a geospatial toolkit for health providers to understand patterns of movement related to health service access. Even though trajectory data mining has been used in animal and ecological studies, public health research has yet to study, collect, and analyze human trajectories data in resource-poor and Global South settings. Given the growing ubiquity and acceptance of *activity spaces* as a conceptual model in the geographical component of the SDoH, the analyses offered in Chapter 6 propose a means to further understand the nuances of individual movement and their implications for overall community health access.

Chapter 3

Methods

3.1 Study setting

In this chapter, I describe a two-phase study design towards understanding community-wide and individual activity movement within rural villages in Malawi. The data collection for this research was conducted at two villages within Lilongwe District, both of which are located in the Central Region of Malawi. To maintain the anonymity of the participants, these villages will now be referred to as Village 1 and Village 2. From Lilongwe, the duration of travel to both villages is highly dependent on weather conditions and season, given that the dirt roads leading to both villages may become washed out or impassable during heavy rains. During the dry season, travel to the village sites from the capital city, Lilongwe, takes an average of 90 minutes, while travel during the wet season could take over two hours. For each village, the average driving time to their respective regional health centers is 20 to 45 minutes. Both villages are predominantly Chewa, the majority ethnicity in Malawi. Villagers generate income through selling surplus crops from subsistence farming, collecting and selling firewood, fishing, and engaging in short-term rural agricultural work outside of their own farms, called *ganyu*. Figure 3.1 shows a map of the locations of both villages in relation to their pertinent health centers.

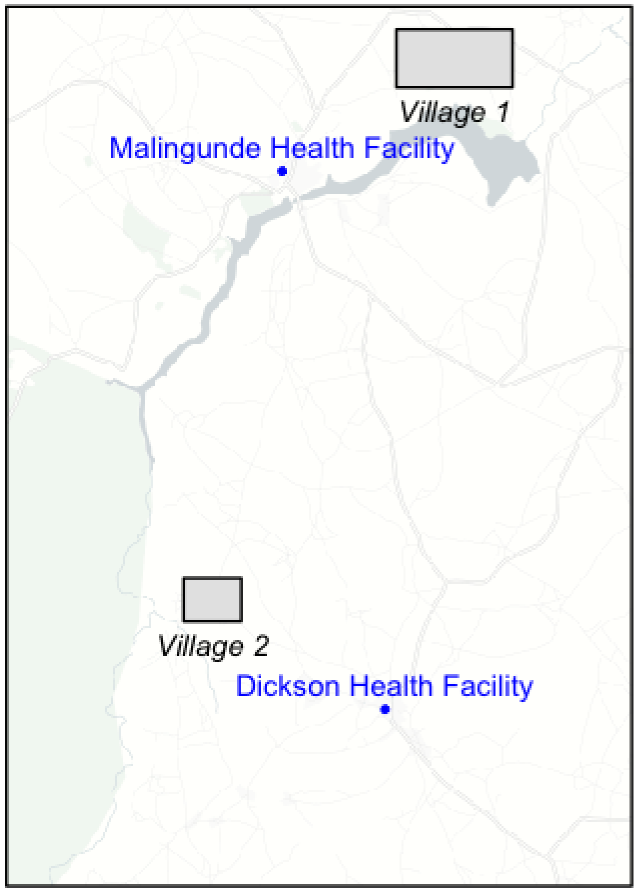


Figure 3.1: Study villages in relation to their closest health facilities

3.2 Partnerships with local researchers and community leaders

This research was made possible through a collaboration between University of California Santa Barbara, University of California Los Angeles, and Partners in Hope in

Lilongwe. In-country and on-site logistics, survey translation services, and hiring of research assistants were organized through the Implementation Science Team at Partners in Hope.

3.3 Phase 1 formative assessment: focus groups and community mapping

3.3.1 Overview

The groundwork for this research was established in this phase. Two rural villages were selected within Lilongwe District (1) for their relatively easy access from the capital, Lilongwe; and (2) established ties from participation in prior research. For the sake of anonymity, I will refer to the villages as Village 1 and 2.

The objective of this formative assessment was to understand the culture, social groups and cliques, daily behavior of village residents, and sociodemographics that may influence a villager's day-to-day movement within their community. The assessment was divided into two activities: focus group discussions (FGDs) separated by gender and guided by an open-ended questionnaire, followed by a community mapping (CM) activity. From these discussions and maps, I also extracted information about spatial understanding and suitable incentives for the subsequent phase of in-depth interviews. Additionally, these focus groups were instrumental in contextualizing and refining the geospatial survey used in Phase 2.

3.3.2 Recruitment for the FGD activities

Six research assistants (RA)–three male and three female—were hired and trained to implement two FGDs and CM activities at Villages 1 and 2. I decided to hire a gender balanced RA team because I planned to recruit the same number of men and women for both focus groups and individual surveys. All RAs were Malawian, and the majority had prior experience working as survey staff at the village sites. The RAs worked in two teams of three, separated by gender, to oversee and implement the research activities for their respective genders. Each team had one RA assigned to the role of *moderator*, and two RAs designated as *scribes*. Moderators were responsible for facilitating the FGD by using the open-ended questionnaire and interfacing with the focus group participants. Scribes took detailed notes of the responses given towards the guided questionnaire and made sure the FGD was proceeding in a timely fashion. As soon as the FGD and CM ended, the two scribes compared the responses they wrote in their individual guided questionnaire and consolidated responses into one document.

Both Villages 1 and 2 had been recruited for their involvement in prior HIV intervention studies. Given the novelty of the research activities, I chose study locations that were already familiar with the study organizers, the research team, and hosting survey staff. With the community leaders' permission, the principal investigator and senior study staff held a meeting with the village leaders, and described the research phases, including overviews of the major themes covered in the planned surveys. This was done a week prior to the planned FGD/CM dates, to allow the community to ask questions, or revoke participation if they felt the need. Once village leaders gave their assent to participate, they were asked to recruit up to 15 men and women who fulfilled the following eligibility criteria:

1. Participants must be 18 years of age or older and
2. have lived in the study village for at least six months in the past 12 months, from the date of the planned interview.

3.3.3 Focus groups and community mapping activity

Focus groups were chosen as the data collection tool for this qualitative assessment because of their ability to contextualize responses derived from individual in-depth surveys, to create community rapport and buy-in for additional research activities, and to help calibrate the design and wording of the survey instruments for Phase 2 [58–60]. The English-version of the guided focus group questionnaire is included as Appendix A.

Community maps and consolidated responses from the guided questionnaire were analyzed for common themes regarding the following topics: 1) common patterns of daily, weekly, and/or yearly villager activity and movement patterns, 2) village ethnic and socioeconomic divisions and groups, 3) differences in how gender affects perspective and outlook on the aforementioned topics. After completing the mapping activity, the paper maps remained with the villagers. RAs used their mobile phones to capture photographs of the paper maps, for further analysis.

3.4 Phase 2 geospatial survey: individual interviews

3.4.1 Overview

The Phase 2 geospatial survey instrument was designed to collect detailed demographic and movement data that can be analyzed for geospatial insights. Phase 1 helped reveal general themes and commonalities regarding a community's overall organizational movement trends. This informed and calibrated the survey instrument used in phase 2 (Appendix B). In Phase 2, 199 villagers across Villages 1 and 2 were individually interviewed (49 women and 50 men from Village 1 and 50 men and women from Village 2).

This survey had two components. The first was a demographic questionnaire designed to collect individual characteristics that affected the respondent's travel behavior. The second was a travel behavior journal that reported detailed travel and movement data for the respondent, starting from where they woke up in the morning and ending at where they returned to rest at the end of the day. This travel behavior journal was linked to a large print-out of a satellite image of their community. The respondent was asked to mark the precise locations to which they traveled regularly, and the path/road they used to reach their destinations. Locations were recorded in temporal order on the map print-out, and linked by these drawn trajectories. The English-version of this survey instrument is shown in Appendix B.

3.4.2 Data collection tools and study equipment

Data was collected and hosted using SurveyCTO through their Android App on tablet computers [99]. All RAs were equipped with and trained on the use of these tablets. The tablets were also set-up to connect to the mobile network with SIM cards loaded with prepaid high speed data. Although both Villages 1 and 2 don't receive mobile reception, the tablets were set-up to automatically upload collected data to the cloud once reception was established.

Additionally, all RAs were given blank copies of village map print-outs, carrying cases for protecting the paper maps from precipitation, fine-tipped felt pens for drawing on paper maps, and randomization chip bags.

3.4.3 Recruitment for the survey activity

For both villages, two male and two female RAs were used to implement these surveys. The male RAs recruited and interviewed male participants only, while the female RAs recruited and interviewed female participants only. For maintaining data consistency and mitigating challenges created by geographic literacy among the participants, RAs were trained to carefully collect geographic data for all activities that require annotation on paper maps.

When the RA team first arrived at a village site, a chief was first located and informed about the arrival of the interview team. The chief should already be aware that the interviews are being conducted, but if needed, the RA team would remind the chief about the purpose of the interview and how the collected data would be used.

When the RA team first arrived at a village site, a chief was first located and informed about the arrival of the interview team. The chief should already be aware that the interviews are being conducted, but if needed, the RA team would remind the chief about the purpose of the interview and how the collected data would be used.

Unlike Phase 1, a shortlist of study participants was not pre-recruited. RAs were trained to sample villagers using the random walk sampling method [100]. Villagers were required to satisty the following eligibility criteria:

1. Participants must be 18 years of age or older and
2. have lived in the study village for at least six months in the past 12 months, from the date of the planned interview.

3.4.4 Random walk sampling

All villagers were sampled using the random walk sampling method [100]. First, each member of the RA team would start at their assigned geolocation. A geolocation table is provided to the RAs as a guide to navigate to their assigned starting position for each day. Each coordinate corresponds to four village areas, the composite of which forms the entirety of the village area. If the RA is somehow not able to navigate to their starting location, they are asked to record their actual location in their *Daily Preamble* (Appendix B). Next, they would proceed with the following steps:

1. Once the RA arrives at their assigned coordinate, they would throw a ball point pen upward and allow it to fall to the ground without any interference. After the pen has dropped, the RA should take careful note of where the pen is pointing. This is the direction that has been randomly selected for the RA to walk in order to identify the first random household to be sampled.

2. The RA would now randomly draw a chip from the randomization chip bag. This number will show which house to sample first, when walking in the direction of the pen's point. For example, if the chip shows the number "three", then they would select the third house they pass for sampling. If multiple adult villagers reside in the household, the RA would gather all who may be eligible who match their gender and are interested in participating, and select a participant by using their randomization chip bag.

3. After selection of the first household, the remaining households to be sampled should be selected by identifying the nearest household 90 degrees to the right of the front door of the last sampled house (when facing the front door from the outside). The next household should be sampled the same way, and so on.

4. If there is ever an instance where there are no more houses to the right of the front door of the just-sampled household, the RA would turn left 90 degrees (when facing the front door from the outside) and sample the next nearest household to the left. After this household is sampled, they would revert back to sampling the household to the right when facing the door.

5. If, after switching left, there are still no more eligible households to sample, then the random walk method can be repeated at the RA's current location by tossing the pen and repeating steps 1-4. Although villagers were never sampled more than once; a given household could theoretically be sampled twice by RAs of different genders.

3.4.5 Survey structure and themes

The survey instrument is composed of the following sections: village census, daily preamble, consent, demographics, and travel behavior journal. Each section is described in detail as follows:

Consent

If an eligible villager is interested in participating, allow them to suggest a comfortable location in their home to conduct the survey. RAs were trained on practicing and have rehearsed a "consent presentation" to distill the core components of the consent language, the purpose of the study activities, and the survey activities. If the villager consents to participating in the study, they would mark the Phase 2 Research Participant Agreement form on the study tablet with a mark of their choosing (to account for differences in literacy among the population).

Village Census

The village census was collected once prior to starting data collection at each village site. During the first day of data collection at a village, the team of RAs would meet with the village chiefs to collect the census information for the village, including the total adult population breakdown between men and women. If the village chiefs did not have the census information, the RAs would call the regional health surveillance assistants (HSAs) for this information.

Daily Preamble

During each day of data collection, each RA was asked to record some observations about the events happening in that village. If there are no events happening, then they would indicate this in the Daily Preamble. For example, if the RAs sample on a market day, this needs to be indicated. This is important because the daily events tell us the villager sampling frame on a given day (e.g. if today is market day, I am likely to sample fewer farmers etc.). Also, if the RA started their random walk at a different location, they were asked to record their starting geocode in their Daily Preamble.

Survey - Demographics

This section gathers information related to the participant's tribal and religious affiliation, age, gender, education, disability and employment status, family structure, and economic well-being.

Survey - Travel Behavior Journal

This is the core survey component that gathers the participant's individual daily movement data. This section is divided into two parts, *Movement within Village* and *Activity Log*.

In the *Movement within Village* section, I contextualize the participant's movement behavior by asking questions about how they geographically define their village. On the provided paper map, the RA asks them to circle the boundary of their census enumeration village (a government designated unit for village communities in Malawi). Since census enumeration villages often encompass multiple smaller communities, the next set of questions asks the participant to outline their own "personal village", if applicable. Finally, I ask the participant to indicate a landmark on their paper map. Here, "landmark" is defined as a location that people living in their census enumeration village would recognize.

In the *Activity Log* section, the participant is asked to detail their travel and movement history on the day prior to the interview. On the provided paper map, they are asked to first indicate the place where they woke up, and the first location they traveled to after waking up. They are then asked to show the trajectory of how they traveled there. In this fashion, the participant sequentially pinpoints the next locations they visited and the trajectories they used, until the final location is marked. For each location visited, the participant reports if they traveled with other people to get there, the mode of transport they used, and the time spent at that place.

In the next section of *Activity Log*, Iask if the reported trajectories and locations are routine in the participant's daily life. Specifically, the participant reports any additional locations they'd normally visit that were not previously reported. Ialso ask them to mark locations that are not part of their normal routine. Finally, Iask them to pinpoint places where they find leisure, recreate, or relax in.

3.5 Human subjects considerations

Ethical approval was obtained from the UC Santa Barbara Human Subjects Committee (Protocol number 6-22-0684) and the Malawi National Health Sciences Research Committee (Protocol number 19/05/2373) Verbal informed consent was obtained from all participants prior to taking part in all research activities, including the Phase 1 FGD/CM activity and Phase 2 surveys. All consent and survey documents were translated to Chechewa, and back-translated to English for language and thematic accuracy.

Chapter 4

Using guided focus groups to understand community-level activity and movement behavior

4.1 Introduction to this analysis

The primary aim of this analysis is to use qualitative and participatory research methodologies to understand how rural Malawians move within their communities and interact with their lived environment. Focus group discussions (FGDs), separated by gender, were organized at two village sites to collect contextual information relevant for future analytical tasks. The discussion responses were annotated by a dedicated notekeeper using a guided open-ended questionnaire response packet. Here, I summarize the results derived from this response packet (Appendix A) and describe how the overall themes abstracted from this qualitative data help frame, inform, and provide input on the subsequent analyses presented in Chapters 5 and 6. Additionally, these focus groups helped fine-tune the questions asked in the Phase 2 geospatial survey described in the prior chapter.

4.2 Results

On February 12th 2019, 13 women and 13 men from Village 1 participated in gender-segregated FGDs. On February 13th 2019, 15 women and 14 men from Village 2 participated in gender-segregated FGDs. For these focus groups, I recruited a wide variety of participants that spanned the eligibility spectrum for this study. The FGDs accomplished two goals. First, the FGDs were helped calibrate the Phase 2 survey instrument. Second, they informed the analytical approaches for the geospatial survey data collected during Phase 2.

The median age of women participating in the Village 1 FGD was 32 years (IQR 30-42), while the median age for men in Village 1 was 40 years (IQR 32-51). The median age for Village 1 women is likely to be higher than reported since three elderly women in the FGD did not know their age. Village 1 women have lived in their village for a median of 27 years (IQR 3-32), while Village 1 men have lived for a median of 39 years (IQR 26-51) in their village. For Village 2 women, the median participant age was 35 years (IQR 30-40) and the median years lived in their community was 31 years (IQR 18-42). Similar to Village 1, the median age of women participants is likely to be higher than reported, because two elderly women participated in the FGD who did not know their age. For Village 2 men, the median participant age and time lived in their community was 32 years (IQR 25-35) and 30 years (IQR 24-35) respectively. Table 4.1 shows the demographic characteristics for the participants.

The detailed results for this analysis are presented below, organized in eight core topics: 1) community history and culture, 2) family structure, 3) village infrastructure, 4) community cliques and seasonal village schedules, 5) village occupations and seasonal work, 6) health and perceptions of people living with HIV (PLHIV), 7) village movement and activity, and 8) participatory mapping. Most of the topics were discussed within the context of assessing village activity and movement. For this purpose, focus groups named salient social and community groups within the village. Next, they discussed in what way these groups may tend to move or gather. Please note that unless notable differences exist between the male and female FGDs, these results report the consolidated responses from both genders disaggregated by village site. Finally, it is important to note that unlike a conventional qualitative analysis, these topics did not emerge from the FGDs as emergent or dominant themes. These topics were predetermined by myself, my mentors, and local collaborators as relevant topics needed to inform the development of the Phase 2 survey and for informing the results of the downstream analyses.

4.2.1 Community history and culture

At both villages, both the male and female FGDs reported that the Chewa tribe founded their communities. At Village 1, the female FGD reported that their village was founded 45 years ago, while the male group was not certain of the year of founding. At Village 2, the female FGD revealed that their village has been in existence for at least 100 years, while the males were again not certain.

In Village 1, the predominant languages spoken are Chichewa, Tumbuka, Yao, and English. Similarly in Village 2, the languages spoken are Chichewa, Ngoni, Chisena, Tumbuka, and Yao. Chichewa is the language spoken by the Chewa tribe, while Tumbuka, Yao, Ngoni, and Chisena are spoken by members of their namesake tribes. The religions represented at Village 1 include Islam, local tribal beliefs, and five types of Christianity: Catholicism, Anglican, Pentecostal, Last Church of God, Presbyterian, and Nazarene. At Village 2, the represented religions are local tribal beliefs, and five types of Christianity:

Presbyterian, Zion, Catholicism, Aneneri, Baptist, and Last Church of God. At both villages, the female FGDs were able to recognize within their communities a greater diversity of tribes, languages, and religions than their male counterparts.

When asked about how their village compares to their neighboring communities, in terms of culture and infrastructure, Village 1 men and women said that their village has a greater diversity of tribes. Although predominantly Chewa, Village 1 also includes individuals from the Yao and Tumbuka tribe. Furthermore, Village 1 is more easily accessed by roads and transport than other villages, has more water pumps, churches, and schools; and serves as the central village for all the neighboring communities. Similarly, Village 2 also serves as the central village for their neighboring communities. Village 2 also has their own clinic, trading center, and school and similarly serves as a central hub for neighboring communities.

4.2.2 Family structure

Polygamy is very common within both villages. Neither the Village 1 male nor female FGDs provided more details on polygamy within their communities. Village 2 verbalized that polygamous relationships happen due to transient work: because many men work seasonally in South Africa or neighboring countries, their wives are encouraged to marry men who remain in their village, for physical protection and financial security.

By extension, it is very common for certain members of a single household to sleep at different homes on different nights. Although polygamy was noted in all FGDs as one reason for this practice, older children are often sent to live with their grandparents at both villages. Specifically, in Village 1, male children sleep separately from their parents and sisters in dwellings called *gowels*.

4.2.3 Village infrastructure

In both villages, neither pathways nor roads are paved. Major roads that traverse through the village are constructed by the government and periodically maintained by work parties within the villages. Smaller pathways and walkways are maintained and created by the villagers, and many of these paths appear organically due to repeated use. During the rainy season, it is common for roads and paths to have potholes, become washed-out, or destroyed altogether.

4.2.4 Community cliques and seasonal village schedules

When asked about social, community, or recreational groups within Village 1, the male and female FGD named the following groups: soccer/football teams, a World Relief Working Group, various religious choir or youth groups, construction or renovation teams, the Nkhonde Bank microlending and rural banking group, a tribal group called the *nyau*, and a men's only mental health and group therapy meeting. The Village 2 responses were the same as Village 1, with the addition of school committees. When asked if people involved in community and social groups tend to live in certain areas of the village, all focus groups said no: members of these groups live scattered around their village, and meet at central locations or venues when needed.

There are a number of regular community events that draw the attendance of many residents. Chief among these are weekly market and church days. In Village 1, market days are held every Friday and Sunday and in Village 2, every Wednesday. Church days are Fridays, Saturday, and Sundays in Village 1, and on Saturdays and Sundays in Village 2. In Village 1, other weekly events include football (soccer) and netball games, dance practice, and Nkhonde Bank meetings. In Village 2, other weekly events mirrored those from Village 1, with the addition of weekly mobile clinics organized by visiting government health officers.

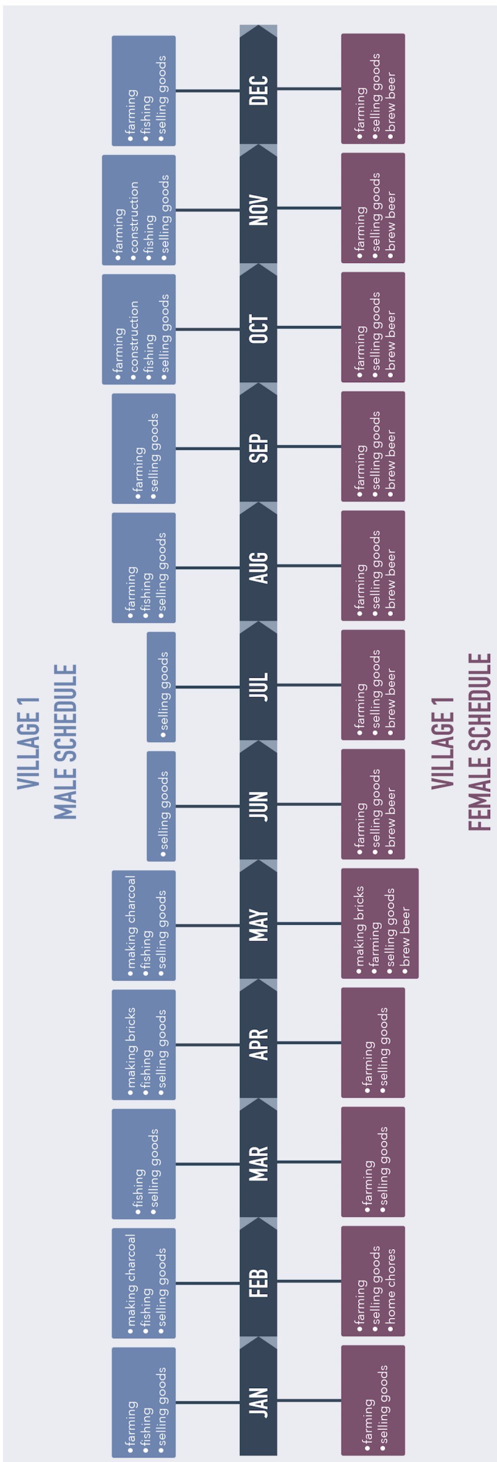
In general, farmers in Malawi sow their crops at the start of the summer rainy season (normally at the end of November) and harvest them at the start of the winter dry season in April or May. When asked if weekly events change between wet/planting and dry/harvest season the answer was “no” from the Village 1 FGDs. In Village 2, both FGDs noted that although weekly events don’t change between seasons, funerals are an exception that would alter the community’s routine. The male FGD in Village 2 also noted that during harvest season, there tends to be more village activity and events compared to planting season, where many villagers are busy with farm activities and other agricultural labor. In addition, the Village 2 male FGD noted that during the harvest season, most villagers have sold their agricultural products and accumulated enough money and food to engage in recreational activities (e.g. traditional dances and festivals).

4.2.5 Village occupations and seasonal work

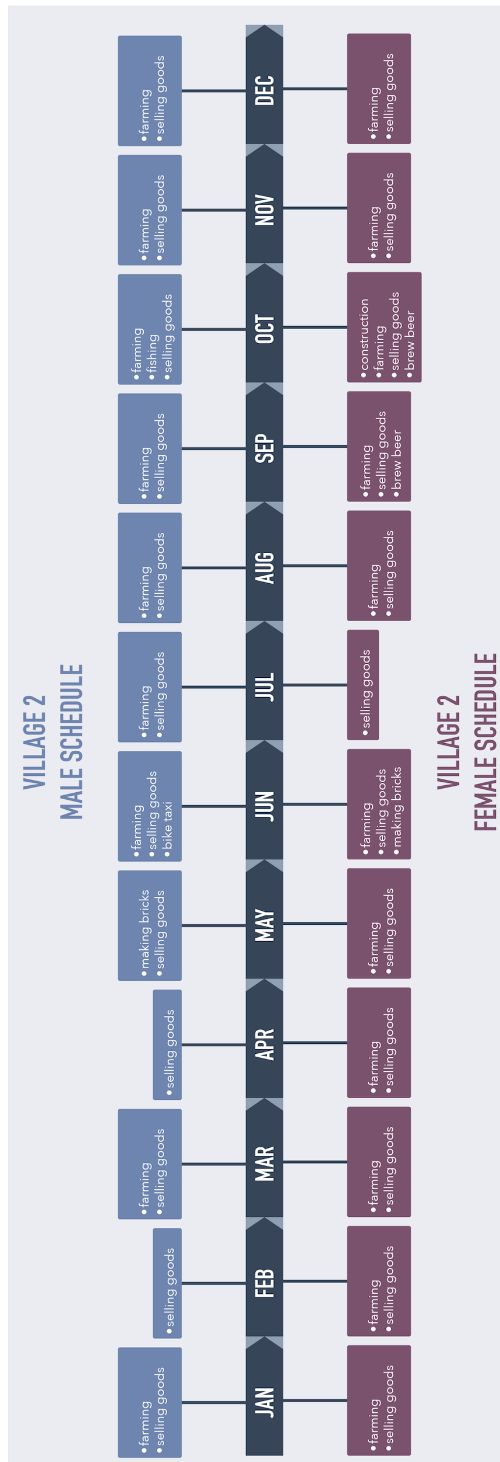
In this portion of the focus groups, the participants were asked to name the two most common occupations for men and women, during every month of the year. A graphic showing these occupational timelines are shown below in Figure 4.1. Next, the participants 1) collated the most popular occupations for each village, and 2) described differences in occupational roles between men and women. These results are shown in two sets of tables (Tables 4.2 and 4.3).

4.2.6 Health and perceptions of people living with HIV

In this section, the participants were asked to describe common health problems faced within their communities. They were also asked about general perceptions of PLHIV



(a) Village 1 yearly employment timeline



(b) Village 2 yearly employment timeline

Figure 4.1: Seasonal occupational schedules (disaggregated by gender) reported by FGD particiapants at Villages 1 and 2

among their village population. In Village 1, common illnesses and health problems include scabies, cholera, HIV/AIDS, tuberculosis, dysentery, and blindness. In Village 2, participants named malaria, HIV/AIDS, dysentery, asthma, tuberculosis, blindness, and high blood pressure as common ailments. Village 1 residents seek care from the Malingunde Health Facility, the Mitundu Regional Hospital, and the Kamuzu Central Hospital in Lilongwe (the national capital). Village 2 residents seek care from the Dickson Health Facility, the Malingunde Health Facility, the Mitundu Regional Hospital, and the Kamuzu Central Hospital in Lilongwe. Both villages are also regularly served by mobile health clinics that provide care and treatment for minor injuries and illnesses. Although most focus groups opined that their basic health needs are met by the facilities close to them, Village 1 women were the exception. They said that their health problems are rarely addressed at their local health clinics, because these facilities are often over encumbered and overworked.

All focus groups voiced that HIV is a problem within their communities. When asked to describe why HIV continues to be an issue within their village, Village 2 women noted that polygamy and having multiple sex partners contributes to community spread of HIV. Village 2 men voiced that HIV contributes to poverty within their village and increases the number of orphans within their community. Village 1 women voiced that HIV is still a persistent problem in their community because some villagers are unwilling to use condoms even after knowing their positive infection status. Also, if the PLHIV is on treatment, it is a considerable challenge for these folks to consistently take their medication, making community-wide viral suppression a challenge. Village 1 men said that PLHIV within their village are stigmatized and segregated, which causes newly infected villagers to occlude their positive status. This leads to challenges for infected individuals to seek care and treatment, which further exacerbates their own illness while driving community spread of HIV.

When asked if PLHIV can be open about their status to members of their village, both FGDs from Village 2 said no: these individuals continue to fear community-wide stigmatization. Village 1 men and women were the exceptions to this, sharing a more welcoming and egalitarian mindset towards PLHIV. They shared that when a PLHIV shares their status, residents within Village 1 are often willing to provide help and support. They said that this change is driven by regular HIV education campaigns by visiting government health officers.

4.2.7 Village movement and activity

In this section, each FGD was asked to brainstorm the most commonly visited locations within their communities, the common modes of transport that are used to visit these locations, and the general frequency of visits for these places. The tables that show these results are Tables 4.4 and 4.5. Common responses between men and women (disaggregated by village) are highlighted in yellow, while differences remain unhighlighted.

These results show that in both villages, although there are common locations that both genders visit, there are also places exclusive to a given gender. In Village 1, locations exclusive to women include the mosque, water pump, forest, graveyard, and school toilets. Locations exclusive to men are bars and family farm plots. In Village 2, only women are reported to visit the school, church, maize mill, graveyard, and river. For Village 2 men, the only exclusive location they commonly visit was the football ground. Further, even though there may be locations that are shared between genders, the mode of transport and frequency of visiting those locations may differ. For example, in Village 1, although both men and women named the water pump as a common location, women visit on a daily basis while men visit every other day.

When asked to describe a typical day for men and women, the FGDs provided a variety of responses. The Village 1 women could not agree on a common descriptor for the schedules of their male counterparts. Some voiced that the men of their village barely work. Others pushed back and said that men provide the raw food ingredients that women are able to use and cook into meals. For themselves, the Village 1 women said that a typical day is busy, including waking up early, doing chores and cooking, and managing the home.

Village 1 men also gave varied responses on how daily schedules differ between genders. For themselves, some admitted that they mostly stay at home. Others indicated that the majority of their day involves fishing and selling their catch, and working to find and provide food for their families. Many Village 1 men agreed that their work is more physically strenuous than their female counterparts. When asked to describe the daily schedules of their female counterparts, Village 1 said that women do most of the household chores and help sell the fish that men catch. Others said that women don't do strenuous work, and mostly stay at home waiting for men to bring food home. The responses from the male and female FGDs from Village 2 largely mirrored those of Village1.

Next, the participants were asked about people living with disabilities within their community. During the FGDs, "people living with disabilities" were defined as: individuals who want to participate in chores and work within their village, but are unable to do so because of injury, age, or disease. All these FGDs voiced that individuals who fall under this definition are facing a disability that impedes their mobility. Some attributes that contribute to poor mobility are old age, blindness, injury, or infectious diseases like tuberculosis or HIV infection. When asked about how these individuals spend their days, most participants voiced that they stay at home. Some will "hang out" at their relatives' homes or at common village areas, like the trading center. Most of the participants said that these individuals rely on their family for support.

4.2.8 Participatory mapping activity

An ancillary component of these FGDs was a participatory mapping activity. The specific instructions for this activity are described in Chapter 3 and Appendix A. In summary, all FGDs were asked to draw a map of their community, noting the village boundaries, roads and pathways, schools and clinics, major geographic features, and other village landmarks. Unfortunately, because of poor image resolution, the map created by Village 2 men was not able to be interpreted and analyzed.

There are two main insights derived from the remaining maps. First, the frequently visited locations drawn on the maps matched the responses provided during the FGDs when asked about village movement and activity (Tables 4.4 and 4.5). Second, it was clear from this mapping activity that group, tribal, or religious membership did not determine where villagers resided within their communities: villagers tended to live dispersed throughout their community, matching the responses provided during the FGDs when asked about community cliques.

4.3 Discussion and conclusion

After considering and collating the findings from the FGDs, this discussion is organized into three takeaways. Each of these takeaways help contextualize the analyses presented in Chapters 5 and 6. Using these FGDs responses, I postulate on whether demographic characteristics or seasonal occupations/employment are likely to impact patterns of activity and movement within rural Malawian villages.

4.3.1 Takeaway 1

The FGD findings helped finalize the Phase 2 geospatial survey

The Phase 1 FGDs were scheduled two weeks ahead of Phase 2, to help refine the questions and survey design of the Phase 2 survey instrument. From the FGD findings, I was able to make revisions and amendments to the demographic section of the questionnaire (Appendix B). Specifically, after reviewing the FGD findings, I finalized the variable choices for a number of demographic variables (e.g. *tribe* and *religious affiliation*). Prior to the FGDs, I wasn't able to complete the selector categories for these variables, given my lack of familiarity with the village sites and their demographic composition. Further, even though the two villages were close in geographic proximity and likely to have similar tribal, religious, and demographic composition, it was unclear if this was the case until I completed the FGDs.

The FGDs helped confirm that the two villages had a very similar demographic composition. Tribally, both villages are predominantly Chewa. With regards to religious composition, both villages are predominantly Christian with a similar denomination profile, along with some adherents to Islam or local/traditional beliefs. Economically, both villages partake in similar economic activity, mostly subsistence farming, fishing, and gig labor. Therefore, instead of curating the survey instrument individually for each village, the FGD results revealed this was not necessary.

4.3.2 Takeaway 2

The FGD results demonstrated that within a village community, income generation and employment shift and change with seasonality

These FGDs revealed that unlike employment trends in industrialized nations, employment varies between men and women, and is deeply dependent on season. These shifts and changes reflect that most rural Malawians engage in some form of subsistence agriculture while bolstering their income with side hustles or gig labor. Figure 4.1 shows that certain work or tasks are shared between genders (e.g. agriculture or selling crops) while some are exclusive to a given gender (e.g. women exclusively brew beer while men solely engage in fishing). Further, some of these occupations are seasonal and restricted to certain months of the year. For example, women predominantly brew and sell beer, but only certain months of the year, depending on the village. In Village 1, brewing beer is a dominant occupation for women between May and December, while in Village 2, it's restricted to September and October. These seasonal employment variations inform how I understand and contextualize the occupation response in the Phase 2 geospatial survey, knowing that the responses I collect will only reflect the work opportunities available during February and March, the months of the data collection activities.

4.3.3 Takeaway 3

The FGD findings revealed that differences in daily activity may be determined by gender

From the responses gathered from these FGDs, I was able to discern that there may be considerable differences in individual activity based on gender. The FGDs revealed that men and women engage in different work over the course of a year (Figure 4.1). In both villages, women are more likely to engage in tasks like brewing beer and chores, while men are more involved in fishing. At the same time, however, men and women share many occupations in agriculture, construction, and mercantilism. These trends are shown in Figure 4.1, where month-by-month occupations and tasks are listed for each gender.

Beyond occupation and employment, Tables 4.4 and 4.5 show that at both village sites, men and women regularly visit different locations within their community. In Village 2, for instance, women listed the school, church, maize mill, graveyard, and river as locations they frequented that men did not name. In Village 2, men named the football ground as a common location that wasn't named by their female counterparts. Further, even when men and women visit the same location, they have a tendency to use different means of transportation to get there, or visit at different frequencies. For instance, both Village 2 men and women frequently visit the water pump, but men reported that they visit three days a week, while women visit on a daily basis.

Although these results demonstrated that there may be two gender-based categories of activity, it is important to note that these differences may, again, vary throughout a given year, depending on season. Both economies of Villages 1 and 2 are heavily dependent on agriculture and fishing, the selling of commodities and resources (e.g. bricks, firewood, and charcoal), and providing labor for village construction and maintenance. As gender roles shift and/or merge according to seasonal work, activity can be expected to change according to season. Therefore, the results and findings from future work in this domain should be interpreted with these contextual shifts in mind, and the specific time of year activity data is collected.

4.3.4 Takeaway 4

The FGD findings helped reveal that differences in daily movement may be determined by gender. However, other demographic factors may not have as big of an impact on daily movement.

These FGDs provided some insight about possible differences in movement among different segments of a village population. First, the FGDs revealed that the study villages had a fairly homogeneous composition of tribal and religious groups, with Christianity (with numerous denominations) being the dominant faith, and the Chewa people being the primary tribe. Neither village is geographically segregated based on demographics: villagers live scattered around their village regardless of tribal affiliation, community clique, or religious membership. Therefore, it is unlikely that religious or tribal differences influence individual movement.

Second, these focus groups did reveal that I am likely to see gender-based differences in trajectories and movement patterns. As discussed in the previous takeaways, although men and women do share occupations, there are a distinct set of tasks assigned to either gender. Additionally, men and women do tend to visit different locations within a village. Consequently, there can be some expectation of gender-based differences when analyzing movement within Villages 1 and 2.

Table 4.1: FGD participant demographics

	female (n = 13)	**male** (n = 13)
Age		
min	19	18
max	46	61
median (IQR)	32 (30, 42)	40 (32, 51)
Years Lived in Village		
min	0.083	17
max	43	53
median (IQR)	27 (3, 32)	39 (26, 51)
Marital Status		
Married	5/13 (38%)	11/13 (85%)
Single	8/13 (62%)	2/13 (15%)
Main Occupation		
Family assistance	4/13 (31%)	0/13 (0%)
Gig labor	5/13 (38%)	0/13 (0%)
Farming	2/13 (15%)	9/13 (69%)
Selling Commodities	2/13 (15%)	0/13 (0%)
Fishing	0/13 (0%)	2/13 (15%)
Construction	0/13 (0%)	1/13 (8%)
Carpentry	0/13 (0%)	1/13 (8%)

(a) Village 1

	female (n = 15)	**male** (n = 14)
Age		
min	22	20
max	53	67
median (IQR)	35 (30, 40)	32 (25, 35)
Years Lived in Village		
min	12	15
max	70	67
median (IQR)	31 (18, 42)	30 (24, 35)
Marital Status		
Married	12/15 (80%)	13/14 (93%)
Single	3/15 (20%)	1/14 (7%)
Main Occupation		
Family assistance	5/15 (33%)	0/14 (0%)
Gig labor	0/15 (0%)	0/14 (0%)
Farming	2/15 (13%)	14/14 (100%)
Selling Commodities	8/15 (53%)	0/14 (0%)
Fishing	0/15 (0%)	0/14 (0%)
Construction	0/15 (0%)	0/14 (0%)
Carpentry	0/15 (0%)	0/14 (0%)

(b) Village 2

Table 4.2: Common occupations and gender roles in Village 1

Occupation	**Who Participates?**	**Role of women**	**Role of men**
selling firewood	women and men	selling and felling wood	selling and felling wood
selling farm products	women	selling crops	
farm work	women and men	all farm work activities	all farm work activities
beer making	women	brewing and selling beer	
laundry and fetching water	women	washing clothes and fetching water	
construction	men		construction
charcoal making	men		burning charcoal and selling charcoal
fishing	men		fishing and selling fish

(a) Village 1 women

Occupation	**Who Participates?**	**Role of women**	**Role of men**
fishing and selling fish	women and men	selling	fishing and selling
selling firewood	women and men	felling firewood, selling firewood within village	felling firewood, selling firewood outside of village
selling farm products	women and men	harvesting and selling crops	harvesting and selling crops
selling livestock	men		buying livestock from farmers and selling livestock
making bricks for construction	women and men	watering soil for making clay; putting clay in mold	digging soil for making clay; putting clay into molds
farm activities	women and men	all farm work	all farm work
selling dry maize	women and men	buy maize from farmers; pound grain into flour; sell flour	buy maize from farmers or harvest maize themselves; sell the maize grains

(b) Village 1 men

Table 4.3: Common occupations and gender roles in Village 2

Occupation	**Who Participates?**	**Role of women**	**Role of men**
farm work	women and men	all farm activities	all farm activities
selling firewood	women and men	felling and selling wood	felling and selling wood
selling rubber trees	women and men	felling and selling rubber trees	felling and selling rubber trees
selling farm products	women and men	selling crops	selling crops
beer brewing	women	brewing and selling beer	
selling cooked food	women and men	cooking and selling food	cooking and selling food

(a) Village 2 women

Occupation	**Who Participates?**	**Role of women**	**Role of men**
selling wooden poles for construction	women and men	cutting poles; selling poles *within* village	cutting poles; selling poles *outside* of village
farm work	women and men	all farm activities	all farm activities
selling firewood	women and men	felling and selling firewood	felling and selling firewood; transporting firewood on bicycles
selling farm produce	women and men	harvesting and selling produce	harvesting and selling produce
making bricks for construction	women and men	watering soil for making clay; tending ovens for baking bricks	digging solid for making clay; putting clay into mold
bicycle taxi	men		transporting people and goods
fishing	men		catching fish
selling mice (common street snack is skewered mice)	women and men	cooking mice	killing mice and selling finished skewers

(b) Village 2 men

Table 4.4: Common locations visited in Village 1

Common locations	Common mode of transport to access	Frequency of visit
church	foot	weekly
mosque	foot	weekly
water pump	foot	daily
trading center	foot, bicycle	daily
forest	foot	daily
graveyard	foot	monthly
football ground	foot	weekly
water dam	foot	daily
school toilets	foot	daily

(a) Village 1 women

Common locations	Common mode of transport to access	Frequency of visit
trading center	foot	daily
water pump	foot	every two days
water dam	foot	daily
church	foot	weekly
bars	foot	daily
family farm plots	foot	daily
football ground	foot	daily

(b) Village 1 men

Table 4.5: Common locations visited in Village 2

Common locations	Common mode of transport to access	Frequency of visit
school	foot	weekly
church	foot	weekly
water pump	foot	daily
maize mill	foot	monthly
trading center	foot	weekly
graveyard	foot	monthly
clinic	foot	weekly
river	foot	daily
family farm plots	foot	daily
forest	foot	daily

(a) Village 2 women

Common locations	Common mode of transport to access	Frequency of visit
trading center	foot	daily
water pump	foot	three days a week
clinic	bicycle	daily
football ground	foot	daily
family farm plots	foot	weekly
forest	bicycle	daily

(b) Village 2 men

Chapter 5

Latent profile analysis of daily time-use composition among villagers in central Malawi

5.1 Introduction to this analysis

The objective of this chapter is to use latent profile analysis (LPA) to identify common activity archetypes of residents among rural villagers in Malawi, using their collected compositional time-use data. In this application, I ask if distinct latent activity archetypes could be identified from 199 individual 24-hour time-use datasets collected from rural residents in Malawi. I formulate three activity classifiers, and assess the latent profile analysis results of each for cultural and contextual suitability. I finish by discussing which classifier was most appropriate for rural Malawi and how the resulting archetypes help inform targeted HIV treatment and prevention care in rural Malawi.

5.2 Analytical methods

5.2.1 Data source

The final version of the survey instrument used (English translation) for this chapter is shown in Appendix B. A sample template of a participant's activity table is shown in question 28f in Appendix B. Each row in the activity table is listed sequentially, starting with the first location the person visited on the day sampled, ending with the last place they visited (usually their home).

5.2.2 Categorizing daily activities of villagers in rural Malawi

I first categorized each activity listed within a participant's activity table according to three classifiers, each with their own set of labels. The first classifier contains two labels: *leisure* and *work*. For this analysis, all activities that involve running errands, chores (house work, cooking), and labor (farm work or construction) are considered work. All other activities are categorized under *leisure*, including resting, attending church or mosque, spending time with friends or family, playing sports, or visiting teahouses/bars.

The second classifier extends the first by including the LoPA and MVPA exertion levels described in Chapter 2. There are four total labels in this classifier: *leisure + sedentary*, *leisure + active*, *work + sedentary*, and *work + active*. Here, *active* and *sedentary* follow the definitions of MVPA and LoPA, respectively.

The third classifier also has four labels: *community*, *leisure*, *employment*, and *chores*. Activities labeled as *community* include tasks related to village governance (e.g. women's group meetings and gatherings of village leaders). *Leisure* tasks include resting/napping, attending church and prayer groups, hanging out with family or friends, playing or attending sports games, and visiting teahouses/bars. Tasks labeled as *employment* include any activity performed for income generation, including selling goods, collecting items for selling, or piece work. Finally, *chores* are all tasks that involve taking care of the home and family, including shopping for food or other necessary items, cooking, chopping wood, or fetching water. Any task that could fall under two or more categories will be given a label that matches the most strenuous activity (e.g. "cooking and rest" would be given a *chore* label, rather than *leisure*).

5.2.3 Statistical analyses of compositional data

Next, three 24-hour compositional profiles were created for each participant, one for each activity classifier. Time-use data is *compositional* or *constrained*, meaning that time spent in one defined activity will inevitably decrease time available for another. Data with this inherent dependency, where the sum of its constituent components add up to a constant, are *compositional* [101]. When analyzing compositional data, studies have demonstrated that results from compositional approaches differed considerably from standard analyses, where the dependency among the constituent time categories are not considered [102–104]. Therefore, standard statistical approaches are not suitable for analyzing time-use data because they do not account for its inherent dependency.

This unique data dependency necessitates the usage of compositional data analysis techniques, prior to implementing latent profile analysis. The aforementioned classifiers help create three sets of time-use compositional profiles for each participant, one for each classifier. For each classifier, I calculate the proportion of time spent in each activity label over a day, and create a *sleep* label for the remaining time spent outside the sum of all activity labels. For example, (applying classifier 2) if a participant reports that they spend five hours in *leisure* and eight hours in *work*, I assume that the remaining 11 hours is spent in *sleep*. This way, the time spent in all activity categories, with the addition of *sleep*, summed up to 24 total hours.

5.2.4 Latent profile analysis on compositional data

Latent class models are statistical tools used for identifying subgroups and populations from large and heterogeneous populations. Originally developed for the analysis of dichotomous indicators, these models can now be used on a variety of variable types, including continuous, rank, count, polytomous, and mixed scale [105]. In this application, I categorize individuals using *latent profile models* on continuous compositional data, a specific type of latent class model. All analyses were conducted using `R` (Version 4.2.1 (2022-06-23)) using the `compositions` package.

First, I account for the dependency within this data by expressing each of these classifiers with their set of isometric log ratios (ILRs), using the ILR transformation function of the `compositions` package in `R`. Specifically, the first classifier (three total labels: sleep, work, leisure) was expressed using two ILRs, the second classifier (five total labels: sleep, leisure + sedentary, leisure + active, work + sedentary, and work + sedentary) was expressed using four ILRs, and the third classifier (five total labels: sleep, community, leisure, employment, and chores) was also expressed using four ILRs. Note that the resulting number of ILR vectors is one less than the total labels for each classifier. These ILR vectors contain the *relative* time-use information sufficient for the analysis of compositional data, and can be used in place of the untransformed vectors of raw time-use proportions.

Next, I use these ILRs as inputs for the compositional latent profile models. Latent profile analysis uses a finite mixture latent modeling approach to derive subgroups on a user-specified set of variables (in this case, relative time-use information) pre-specified number of profiles. Further, these models are used to identify subgroups that are homogenous *within* groups and heterogeneous *between* groups, with respect to these chosen variables.

Finally, although there isn't an accepted standard for choosing the number of estimated profiles while implementing latent profile models, researchers tend to use the following measures to determine model fit [105]:

Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC)

These indicators consider the model's goodness-of-fit and parsimony, where lower values of both indicate that a model is a better fit for the chosen data. As a cautionary note, AIC and BIC comparisons are used within classifiers only, or in other words, within model comparisons that consider the same set of dependent variables. Given that each classifier uses a different set of activity labels, AIC and BIC comparisons are made between profile estimations within each classifier.

Bootstrap Likelihood Ratio Test (BLRT)

This test uses a bootstrap resampling method to approximate the p-value of a generalized likelihood ratio test comparing a k-profile model with a $k-1$ profile model. In this case, a p-value less than 0.05 indicates that the k-profile model provides a significantly better fit to the data than a $k-1$ profile model.

Entropy

This value ranges from 0 to 1 for each profile solution. Higher entropy values show a higher degree of classification certainty. After choosing the best model for each classifier, each participant was assigned a profile based on their maximum posterior probability of being in that profile [106].

Pairwise tests

Z-tests, analysis of variance, and Student's t-test were used to compare differences in demographic characteristics between profiles within the final models.

5.3 Results

In this analysis, I analyze 199 individual time-use datasets collected from the study participants. Table 5.1 shows the model indicators of one to five latent profile estimations for each classifier used to categorize the participant's daily activity data, for a total of 15 models. For classifier 1, the AIC and BIC values vary between the number of profiles estimated. Entropy values trend downwards as the number of profiles increase. Similarly, the BLRT values vary between models, and tend to favor those with an even number of profiles. Ultimately, the two-profile estimation is the best fit model for classifier 1, given its notably high entropy value, significant BLRT indicator ($p < 0.01$), and comparatively low AIC and BIC value.

For classifier 2, the three-profile solution is the most suitable model, given a significant BLRT indicator ($p < 0.01$) and high entropy (0.98). Furthermore, the AIC and BIC values are comparatively low, although higher-profile solutions display lower AIC/BIC values. The three-profile solution is still the most appropriate model classifier 2, given that with a high number of profiles, the individual counts within each subclass would get smaller. Given the low sample size of this analysis, I aim to have each profile have at least 10 members.

With a similar consideration of these goodness of fit indicators, the three-profile solution is the most appropriate model for classifier 3. This model best balances entropy (0.97), BLRT ($p < 0.01$), and comparatively low AIC/BIC individuals, while maintaining 10 or more individuals in each profile group.

Table 5.1: LPA model indicators for each activity classifier

Activity Classifier	Profiles	AIC	BIC	Entropy	BLRT (p-value)
1	1	1149.15	1162.32	1.00	-
	2	1108.42	1131.47	0.87	0.01
	3	1113.74	1146.68	0.41	0.58
	4	1075.88	1118.69	0.55	0.01
	5	1081.49	1134.18	0.50	0.75
2	1	2208.89	2235.23	1.00	-
	2	2104.23	2147.05	0.95	0.01
	3	1902.00	1961.28	0.98	0.01
	4	1833.98	1909.73	0.95	0.01
	5	1814.92	1907.14	0.82	0.01
3	1	1656.39	1682.74	1.00	-
	2	1572.48	1615.29	0.89	0.01
	3	1424.18	1483.46	0.97	0.01
	4	1325.28	1401.03	0.97	0.01
	5	1242.10	1334.32	0.96	0.01

Figures 5.1, 5.2, and 5.3 show the daily time-use composition for the selected profile-estimation models for each classifier. Tables 5.2, 5.3, and 5.4 show the demographic composition for the profiles abstracted from each classifier. For classifier 1 (Figure 5.1), the two estimated profiles demonstrate similar profiles of daily time-use, with analogous proportions spent in work, leisure, and sleep. Using a Pearson's Chi-squared test, profile 1 is composed of significantly more men than women ($\chi^2 = 3.9415$, $df = 1$, $p - value < 0.05$). Significant differences are not observed in the other demographic indicators (Table 5.2).

In the three-profile estimation model for classifier 2 (Figure 5.2), more distinct time-use differences are observed between the three profiles, although distinctions are still not pronounced. Profile 1 is characterized by less time in active work; while profile 2 has a higher proportion of time spent in sleep, less time in sedentary leisure activities, and more time working actively. A Pearson's Chi-squared test reveals that the gender composition is significantly different among the three profiles estimated for classifier 2 ($\chi^2 = 8.1474$,

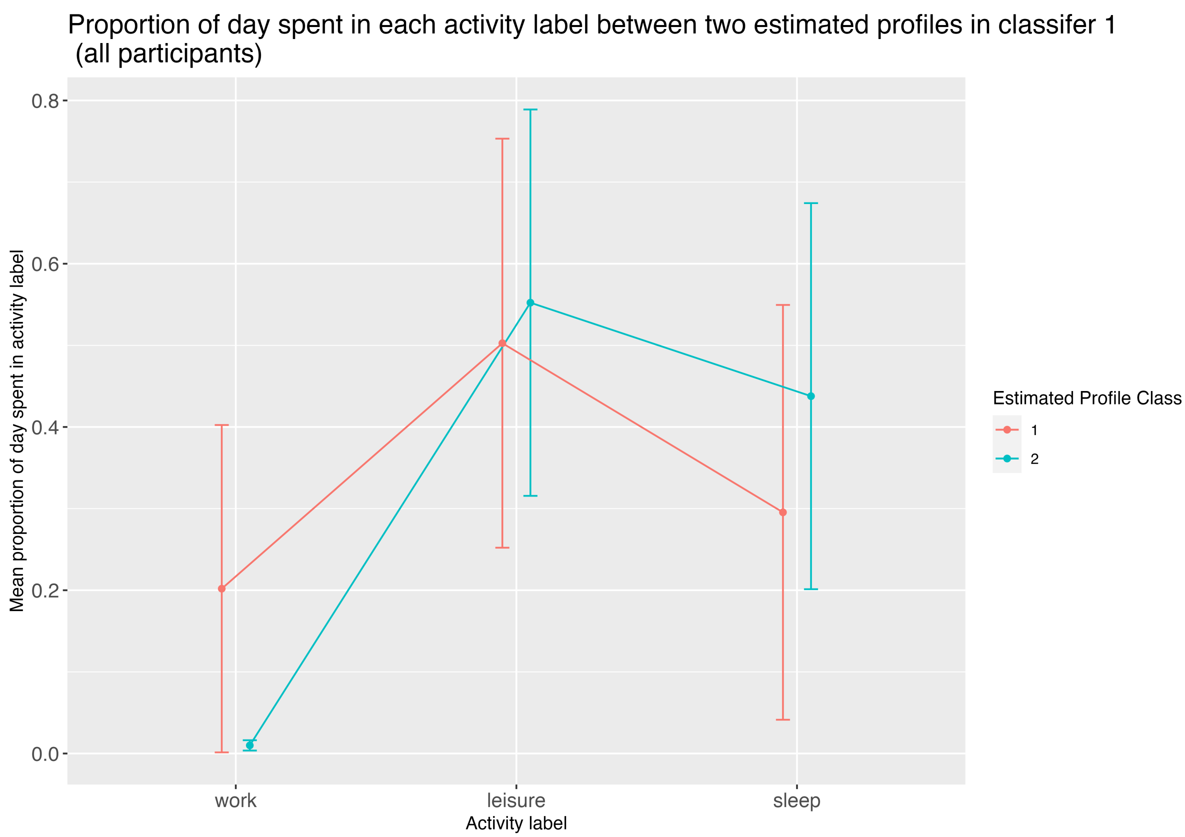


Figure 5.1: Time-use proportions between two estimated profiles (classifier 1)

$df = 2$, $p - value < 0.05$). Profile 2 has significantly more women, while profile 3 has the most men. Significant differences also exist in the composition of occupations among the profiles ($\chi^2 = 29.407$, $df = 10$, $p - value < 0.05$). Both profiles 2 and 3 have a high proportion of participants working in subsistence farming, and profile 3 has the highest proportion of individuals with "other" selected as their occupation. Many participants in this category sell commodities like firewood, livestock, and rubber trees. Finally, profile 2 has the highest percentage of workers in food and vegetable stalls (Table 5.3).

Finally, the three-profile estimation for classifier 3 (Figure 5.3) creates the most distinct time-use profile among its estimated subclasses. Based on these distinct time-use differences, these profiles are named "Balanced" (Profile 1), "Chillers" (Profile 2), and

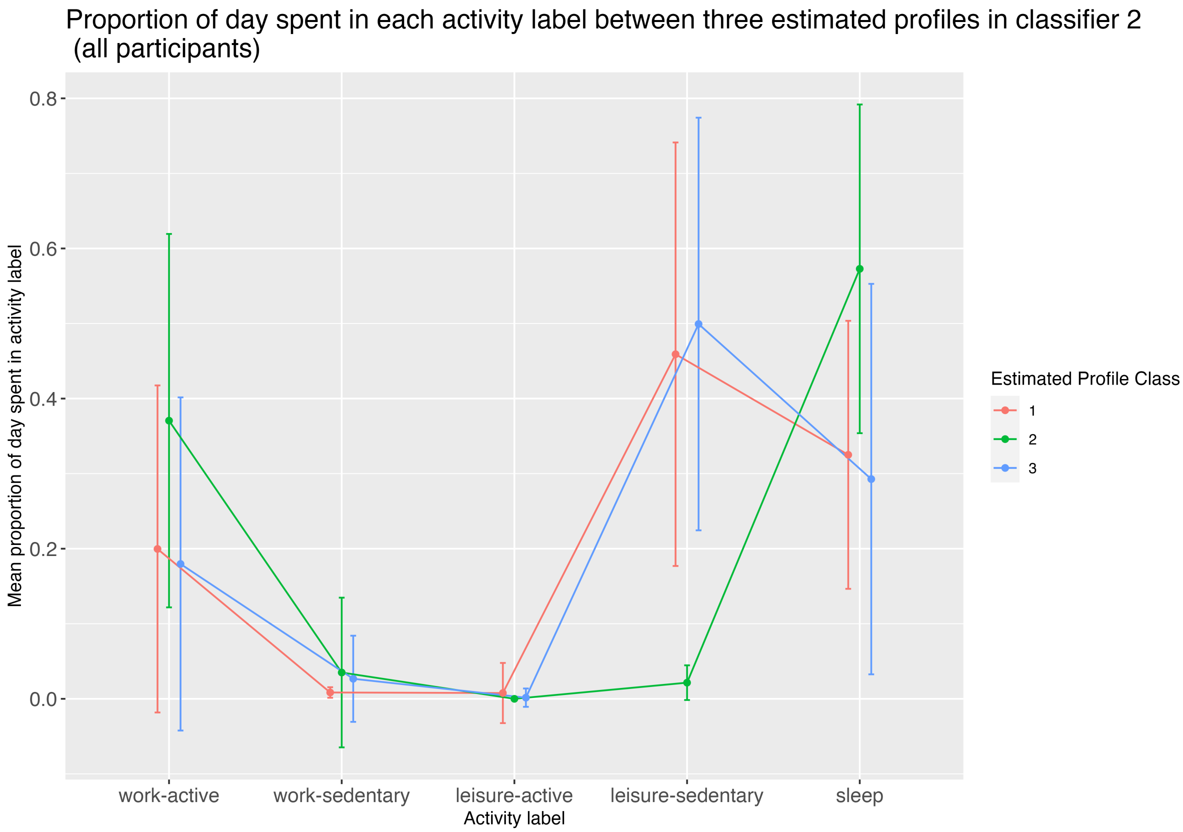


Figure 5.2: Time-use proportions between three estimated profiles (classifier 2)

"Doers" (Profile 3). Of these three profiles, "Doers" spend the most time in chores while "Chillers" do the least. In leisure, "Doers" spend the least time in this activity, while "Chillers" spend the most. A Pearson's Chi-squared test revealed that the gender composition is significantly different among these three profiles ($\chi^2 = 15.757$, $df = 2$, $p - value < 0.01$). "Chillers" have significantly more men, while "Doers" have the most women. The "Balanced" profile has an even balance between genders. Significant differences also exist in the composition of occupations among these profiles ($\chi^2 = 26.892$, $df = 10$, $p - value < 0.01$). The "Balanced" villagers and "Chillers" have the highest proportion of participants working in subsistence farming. "Chillers" also have the highest proportion of individuals working as bicycle taxi drivers (Table 5.4).

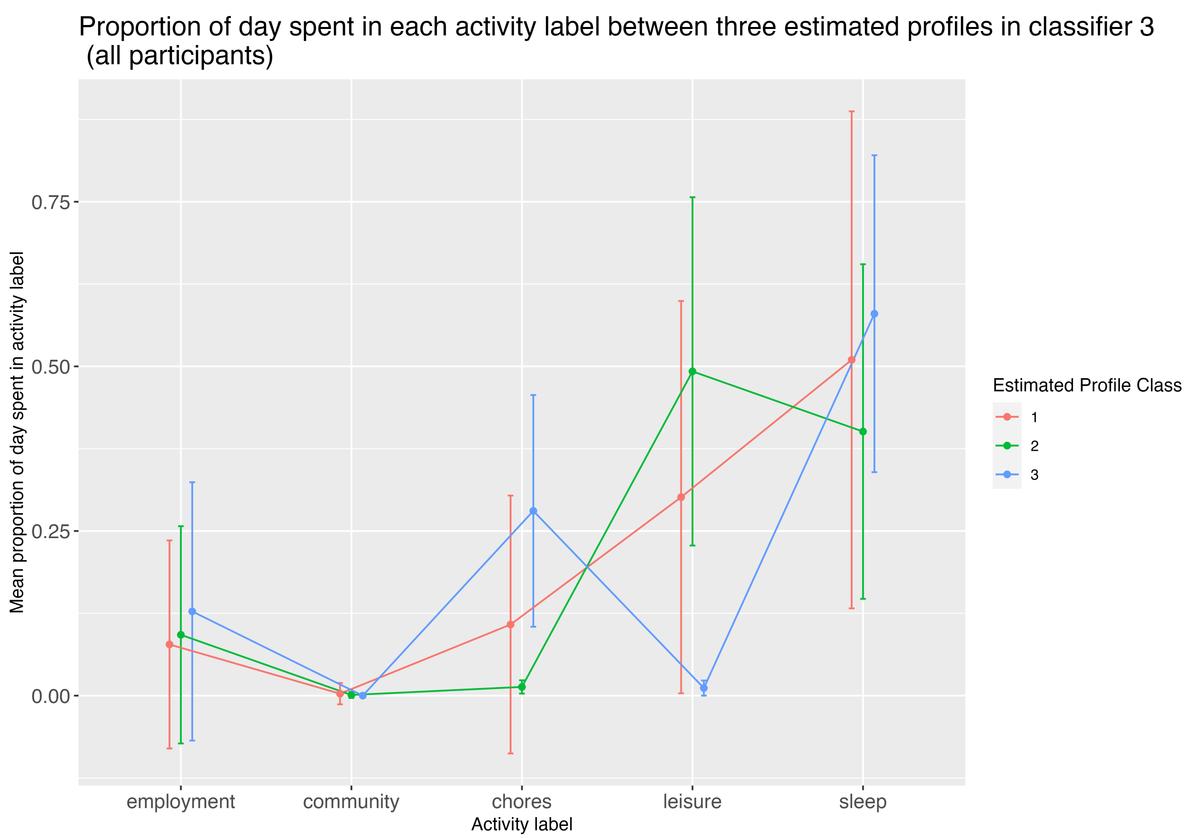


Figure 5.3: Time-use proportions between three estimated profiles (classifier 3)

Table 5.2: In *classifier 1* the participants' activities were labeled as work, leisure, or sleep. The two-profile estimation model was chosen as the most suitable LPA model.

	Total *(n = 199)*		**Profile 1** *(n = 175)*		**Profile 2** *(n = 24)*			
Variables	n (%)	mean (SD)	n (%)	mean (SD)	n (%)	mean (SD)	χ^2 or F	p
Age (years)		38.6 (19.9)		38.2 (20.1)		41.4 (18.9)	0.56	0.45
Gender							3.9	0.047*
Female	99 (49.7)		82 (46.9)		17 (70.8)			
Male	100 (50.3)		93 (53.1)		7 (29.2)			
Are you currently married?							2.2	0.14
No	33 (16.6)		26 (14.9)		7 (29.2)			
Yes	166 (83.4)		149 (85.1)		17 (70.8)			
Have you ever attended formal education?							0.65	0.42
No	33 (16.6)		26 (14.9)		7 (29.2)			
Yes	166 (83.4)		149 (85.1)		17 (70.8)			
How many children currently live with you?		2.97 (1.77)		3.04 (1.82)		2.54 (1.35)	1.65	0.20
How many people currently live in your household?		4.92 (1.87)		4.95 (1.93)		4.67 (1.40)	0.50	0.48

Age of youngest child (years)	5.3 (4.9)	5.4 (5.0)	4.6 (3.8)	0.46	0.5
Current working status				2.72	0.44
Business owner	150 (75.4)	134 (76.6)	16 (66.7)		
Working formally	5 (2.5)	5 (2.9)	0 (0.0)		
Working informally	14 (7.0)	11 (6.3)	3 (12.5)		
Not working	30 (15.1)	25 (14.3)	5 (20.8)		
Primary occupation				2.70	0.75
Unemployed	4 (2.0)	4 (2.3)	0 (0.0)		
Informal labor	30 (15.1)	27 (15.4)	3 (12.5)		
Subsistence farming	93 (46.7)	83 (47.4)	10 (41.7)		
Food and vegetable stalls	10 (5.0)	9 (5.1)	1 (4.2)		
Bicycle taxi	3 (1.5)	2 (1.1)	1 (4.2)		
Other	59 (29.6)	50 (2.86)	9 (37.5)		
Do you currently have a disability that affects your mobility?				1.9e−30	1
No	156 (78.4)	137 (78.3)	19 (72.9)		
Yes	43 (21.6)	38 (21.7)	5 (20.8)		
With your disability, what is your mobility level?				0.92	0.63
IM	156 (78.4)	137 (78.3)	19 (72.9)		

IMA	43 (21.6)	38 (21.7)	5 (20.8)
NIM	43 (21.6)	38 (21.7)	5 (20.8)

Notes: Pearson's chi-squared tests (χ^2) were used to assess differences in categorical membership counts while one-way analyses of variance ($F - statistic$)were used to assess differences in means between the profiles. Values with * are statistically significant at $p < 0.05$.

Abbreviations: IM, can move independently without assistance; IMA, can move independently with assistance (with a cane, walker, etc.); NIM: cannot move independently even with assistance; SD, standard deviation; χ^2, chi-square.

Table 5.3: In *classifier 2* the participants' activities were labeled as work-active, work-sedentary, leisure-active, leisure-sedentary, or sleep. The three-profile estimation model was chosen as the most suitable LPA model for this label grouping.

	Profile 1 *(n = 27)*		**Profile 2** *(n = 11)*		**Profile 3** *(n = 161)*			
Variables	n (%)	mean (SD)	n (%)	mean (SD)	n (%)	mean (SD)	χ^2 or F	p
Age (years)		36.9 (16.5)		39.6 (21.8)		38.8 (20.4)	0.15	0.70
Gender							8.15	0.017*
Female	14 (51.9)		10 (90.9)		75 (46.6)			
Male	13 (48.1)		1 (9.09)		86 (53.4)			
Are you currently married?							0.68	0.71
No	3 (11.1)		2 (18.2)		28 (17.4)			
Yes	24 (88.9)		9 (81.8)		133 (82.6)			
Have you ever attended formal education?							3.4	0.18
No	8 (29.6)		5 (45.5)		36 (22.4)			
Yes	19 (70.4)		6 (54.5)		125 (77.6)			
How many children currently live with you?		2.6 (1.53)		3.45 (1.37)		3.0 (1.83)	0.63	0.43
How many people currently live in your household?		4.59 (1.58)		5.45 (1.86)		4.94 (1.92)	0.40	0.53

Age of youngest child (years)		5.3 (4.8)		4.1 (2.1)		5.4 (5.0)	0.11	0.75
Current working status							7.6	0.27
Business owner	21 (77.8)		5 (45.5)		124 (77.0)			
Working formally	1 (3.7)		0 (0.0)		4 (2.5)			
Working informally	2 (7.4)		2 (18.2)		10 (6.2)			
Not working	3 (11.0)		4 (36.4)		23 (14.0)			
Primary occupation							29.4	$< 0.01^{\dagger}$
Unemployed	0 (0.0)		0 (0.0)		4 (2.48)			
Informal labor	2 (7.41)		0 (0.0)		22 (13.7)			
Subsistence farming	10 (37.0)		6 (54.5)		80 (49.7)			
Food and vegetable stalls	4 (14.8)		3 (27.3)		6 (3.73)			
Bicycle taxi	2 (7.4)		0 (0.0)		1 (0.62)			
Other	9 (33.3)		2 (18.2)		48 (29.8)			
Do you currently have a disability that affects your mobility?							0.28	0.87
No	22 (81.5)		9 (81.8)		125 (77.6)			
Yes	5 (18.5)		2 (18.2)		36 (22.4)			
With your disability, what is your mobility level?							4.2	0.38

IM	3 (11.1)	2 (18.2)	32 (19.9)
IMA	1 (3.7)	0 (0.0)	3 (1.86)
NIM	1 (3.7)	0 (0.0)	1 (0.62)

Abbreviations: IM, can move independently without assistance; IMA, can move independently with assistance (with a cane, walker, etc.); NIM: cannot move independently even with assistance; SD, standard deviation; χ^2, chi-square.
Notes: Pearson's chi-squared tests (χ^2) were used to assess differences in categorical membership counts while one-way analyses of variance ($F - statistic$)were used to assess differences in means between the profiles. Values with * and † are statistically significant at $p < 0.05$ and $p < 0.01$, respectively.

Table 5.4: In *classifier 3* the participants' activities were labeled as work, community, employment, or sleep. The three-profile estimation model was chosen as the most suitable LPA model for this label grouping.

	Profile 1 *(n = 168)*		**Profile 2** *(n = 21)*		**Profile 3** *(n = 10)*			
Variables	n (%)	mean (SD)	n (%)	mean (SD)	n (%)	mean (SD)	χ^2 or F	p
Age (years)		38.9 (19.9)		36.0 (19.4)		38.1 (22.7)	0.18	0.67
Gender							15.76	$< 0.01^{\dagger}$
Female	84 (0.5)		5 (23.8)		10 (100.0)			
Male	84 (0.5)		16 (76.2)		0 (0.0)			
Are you currently married?							0.17	0.92
No	28 (16.7)		3 (14.3)		2 (0.2)			
Yes	140 (83.3)		18 (85.7)		8 (0.8)			
Have you ever attended formal education?							1.63	0.44
No	39 (23.2)		6 (28.6)		4 (0.4)			
Yes	129 (76.8)		15 (71.4)		6 (0.6)			
How many children currently live with you?		3.0 (1.8)		2.2 (1.6)		3.5 (1.4)	0.097	0.76
How many people currently live in your household?		4.9 (1.9)		4.7 (1.7)		5.4 (1.9)	0.1	0.75

Age of youngest child (years)	5.6 (5.1)	4.6 (3.9)	3.73 (2.13)	1.7	0.19
Current working status				11.5	0.74
Business owner	129 (76.8)	17 (81.0)	4 (40.0)		
Working formally	4 (2.38)	1 (4.76)	0 (0.0)		
Working informally	12 (7.14)	1 (4.76)	1 (10.0)		
Not working	23 (13.7)	2 (9.52)	5 (50.0)		
Primary occupation				26.9	$< 0.01^{\dagger}$
Unemployed	4 (2.38)	0 (0.0)	0 (0.0)		
Informal labor	25 (14.9)	0 (0.0)	5 (50.0)		
Subsistence farming	81 (48.2)	11 (52.4)	1 (10.0)		
Food and vegetable stalls	8 (4.76)	2 (9.52)	0 (0.0)		
Bicycle taxi	1 (0.6)	2 (9.52)	0 (0.0)		
Other	49 (29.2)	6 (28.6)	4 (40.0)		
Do you currently have a disability that affects your mobility?				1.1	0.58
No	131 (78.5)	18 (85.7)	7 (70.0)		
Yes	37 (22.0)	3 (14.3)	3 (30.0)		
With your disability, what is your mobility level?				2.7	0.61

IM	32 (19.0)	3 (14.3)	2 (20.0)
IMA	3 (1.79)	0 (0.0)	1 (10.0)
NIM	2 (1.19)	0 (0.0)	0 (0.0)

Abbreviations: IM, can move independently without assistance; IMA, can move independently with assistance (with a cane, walker, etc.); NIM: cannot move independently even with assistance; SD, standard deviation; χ^2, chi-square.
Notes: Pearson's chi-squared tests (χ^2) were used to assess differences in categorical membership counts while one-way analyses of variance ($F - statistic$)were used to assess differences in means between the profiles. Values with * and † are statistically significant at $p < 0.05$ and $p < 0.01$, respectively.

5.4 Discussion and conclusion

At the time of writing this dissertation, this is the first study to explore latent activity profiles based on 24-hour time-use compositions among rural villagers in central Malawi. More generally, this is the first LPA modeling activity to be applied on compositional time-use data within a Global South context. Furthermore, through the usage of different activity label classifiers, this is the first study to use activity labels that deviated from the context of energy expenditure, and examined individual activity as a function of community life.

The third activity classifier (organizing activities into chores, employment, community, leisure, and sleep) reflects the *function* of the activity with respect to village life, rather than the exertion level needed to engage in the activity. Using this classifier, the LPA algorithm was able to best distinguish differences in daily time-use, creating the most distinct time-use differences with its three-profile model (Figure 5.3). Activity and demographic differences are also observed in the estimated profiles from classifiers 1 and 2, which both use energy-expenditure activity classifiers for categorizing the participants' activities. However, compared with classifiers 1 and 2, the LPA model using classifier 3 is able to estimate the most distinct profiles with respect to time-use and demographic differences. These profiles are named "Doers", "Chillers", and "Balanced".

All three profiles from the third activity classifier presented with similar time-use compositions in sleep, community, and employment activities. However, compared to other profiles, "Doers" are characterized by spending the most time on chores and least time on leisure activities on a daily basis. In comparison, "Chillers" spend the most time in leisure and least time performing chores. "Balanced" villagers neither spend the most time nor least time in chores or leisure. However, it is notable that most of the study population ($n = 168$) are categorized into the "Balanced" profile, leaving much fewer counts for “Doers” (n = 10) and “Chillers” (n = 21).

A possible explanation for these membership counts may be explained by our findings in Chapter 4. The demographic composition of these three profiles, in addition to their count membership, is consistent with the insights collated from the focus group discussions implemented in Chapter 4. These focus groups revealed that there appear to be two core activity archetypes among the villagers, divided along gender lines. In the focus groups, I found that men and women occupy different recreation and employment spheres, and I expected that these differences be extracted through the LPA models. However, the focus groups also revealed that although some daily tasks were more likely to see participation by a certain gender (i.e. home chores are largely completed by women and men were more likely to participate in leisure activities), many activities were also shared by both men and women. For example, both men and women would collaborate on the village infrastructure projects, daily agricultural work, and community leadership. This would explain the high membership count in the “Balanced” profile in classifier 3, suggesting that the hypothesized gender roles are not as distinct as expected in the study population. However, these profile distinctions could become more visible given a larger sample size, which presents a core limitation to this research activity.

5.4.1 Using alternative classifiers for labeling activity to derive latent classes

In some health and disease studies an individual’s activity data is commonly used to track energy expenditure to study the relationship between physical activity and various health outcomes (e.g. becoming overweight, developing obesity and/or heart disease, etc.). In a latent class analysis of a cohort of Danish workers from three employment sectors—cleaning, transport, and manufacturing—a participant’s activity was labeled according to nine categories: sedentary, standing still, LoPA, MVPA at work and leisure, and time in bed [107]. These categories are delineated based on energy-expenditure, and are often used in nutritional and adiposity research applications.

The utility of an energy expenditure-based labeling scheme is less applicable in a Global South context, where village life already involves a high baseline level of physical activity. In this chapter, the analysis of rural time-use data shows that both male and female villagers spend a large portion of their day engaging in physical chores, agricultural work, and other active employment modalities. In other words, to abstract differences in activity trends among a village population, differences in physical exertion is not sufficient to distinguish different activity profiles.

Instead, the classifier that produced the most distinct profiles categorized activities based on its function within a community, rather than its degree of energy expenditure. Why is this significant for health providers, and in particular, the provision of HIV/AIDS treatment and testing within rural villages? Within the field of HIV implementation science in Global South contexts, it is important to identify individual characteristics and behaviors that may influence an individual to enter the testing and treatment cascade. Therefore, activity profiles based on energy expenditure are less applicable in this context. Instead, important considerations for HIV testing and treatment include understanding specific behaviors that allow service providers to target individuals for recruitment. The LPA model generated from the third activity classifier showed that two distinct profiles, "Doers" and "Chillers", are largely distinguished by gender. Our results showed that "Doers" are mostly women and engage in more chores, while "Chillers" are mostly men and engage in more leisure activities. Therefore, testing and treatment protocols that wish to target a specific gender for recruitment could be informed based on these activity profiles: men could be recruited at leisure locations while women could be targeted in places where chores are performed.

5.4.2 Strengths and limitations

The core strength of this study is its unique application within a rural Global Health context, and its consideration of activity labeling beyond the energy expenditure framework. Furthermore, the use of compositional data analysis permitted the dependent nature of the 24-hour time-use data to be respected and considered.

One limitation of this study was insufficient sample size. This resulted in low membership counts for a number of estimated profiles. When demographic differences were observed between profiles, the low membership counts created uncertainty in these differences. Additionally, due to the self-reported nature of the data collected, some collected data was unreliable for certain individuals, resulting in time-use reporting that didn't sum up to 24 hours, or sleep times being under or over-reported. Finally, the external validity of these findings is unknown, given the lack of activity behavior studies in rural Global South contexts. To address these limitations, future studies could consider implementing time-use survey methodologies incorporating wearable activity trackers to accurately capture villagers' activity information. Also, given the high proportion of individuals residing in rural areas across the Global South [6], future time-use studies should be implemented in other villages across Malawi and other LDC, to better understand activity archetypes within these rural communities.

Chapter 6

Using dynamic time warping to compare villager movement trajectories from rural Malawi

6.1 Introduction to this analysis

In this final analysis, I present another applied clustering method for analyzing individual geospatial data. Concurrent with collecting daily activity tables from the participants, movement trajectories were also recorded. The aim of this chapter is to cluster similar trajectories using dynamic time warping, while comparing salient demographic profiles of each cluster. In this analytical activity, I asked the following questions: What are the differences I can observe between the spatial features of each cluster? What are the demographic differences, if any, that can be observed from dynamic time warping clustering? How might this clustering technique be useful for rural health providers serving these two village communities? Before I introduce the usage of dynamic time warping (DTW) in this application, I will first introduce the DTW similarity measure, and explain why it was chosen to cluster the collected trajectories in this study. Please refer to Chapter 2 for a detailed background of various similarity measures used when comparing and analyzing trajectories.

6.1.1 Partitional clustering using the DTW algorithm

The DTW algorithm searches recursively through all point combinations between two trajectories, to calculate the distance between them. These collated distances ultimately represent the degree of similarity between two trajectories. Given two trajectories T_A and T_B of length m and n respectively, an m-by-n distance matrix is first created, where element (i, j) represents the distance between i^{th} element in T_A and j^{th} element in T_B. In many implementations, this distance measure can be specified by the user (e.g. Gaussian Alignment Kernel Distance, etc). Table 6.2 describes in greater detail the specific distance function employed in this chapter's analysis.

The next step is to find the best alignment between T_A and T_B. With this distance matrix, a path is drawn through it that minimizes the cumulative distance between the two sets of points. This path, called the *warping path* or *warping cost*, begins at the first element of each trajectory (the bottom left hand corner of the matrix) and ends at the last element of each trajectory (the upper right hand corner of the matrix). The warping path can only traverse the matrix contiguously (by moving forward one step at a time) and monotonically (by only moving forward and not backward). Usually, the warping path follows the diagonal of the distance matrix.

Finally, to cluster the trajectory data based on the warping cost calculated by the DTW algorithm, one needs to employ hierarchical, partitional, or fuzzy clustering [108]. Hierarchical clustering attempts to impose a hierarchy of groups, in which clusters are created by merging data from the lowest level to the next, such that ordered groupings are obtained. Partitional and fuzzy clustering both aim to create partitions in the data. While partitional clustering creates distinct partitions, fuzzy clustering creates a *soft* partition in which cluster membership has an associated probability. Although hierarchical clustering creates distinct groups after "cutting" the ordered tree structure at a given point, it is algorithmically complex compared to partitional clustering [108]. Given this, I chose to employ partitional clustering in this application.

Partitional clustering begins by initializing k centroids, by randomly choosing k objects, or trajectories, from the data to initiate k clusters. The distances between the centroids and all trajectories in the data are then calculated, and each trajectory is assigned to its closest centroid, based on the DTW distance measure. A prototyping function is then used to update the positional value of the cluster's centroid. The DTW distance is again calculated between all trajectories and updated centroids. This recalculation may result in trajectories needing to change clusters based on updated DTW distance values. This process repeats for a given number of iterations, or, until trajectories no longer change clusters.

6.1.2 Using radius of gyration to characterize individual and community mobility

The radius of gyration (RG) is a metric that assesses the overall spatial extent for a defined set of locations or trajectories. In conjunction with DTW clustering, the RG measure helps quantify the spatial extent of the trajectories clustered. In this analysis, this measure serves as a comparable and quantifiable metric to define the spatial extent of each cluster generated.

The RG is defined as the accumulated distances of deviation from the centroid of all the recorded locations of each participants' spatial path or trajectory. In numerous analyses and studies on human mobility and movement, the RG has been used to approximate the spatial dispersion of an individual's activity space [109–113]. For a given individual u's trajectory, its RG is defined as follows [110, 114]:

$$r_g(u) = \sqrt{\frac{1}{n_u}\sum_{i=1}^{n_u} dist(r_i(u) - r_{cm}(u))^2}$$

where $r_i(u)$ represents the nu positions recorded for u's trajectory, and $r_{cm}(u)$ is a constant representing the center of mass of u's trajectory.

6.2 Analytical methods

6.2.1 Trajectory data source

Appendix C shows a paper map with hand-recorded trajectories from one of the study participants. To briefly summarize this process, Participants were asked to report two main types of trajectories: 1) The daily and/or usual trajectories for day-to-day movement and 2) trajectories to visit places of leisure or fun. These two sets of trajectories served as the analytical data source for this clustering analysis. These trajectories correspond with the activity tables used in the latent profile analysis presented in Chapter 5. Specifically, for each row listed in a participant's activity table, there is a corresponding trajectory reported on a paper map.

6.2.2 Digitization of trajectories and data cleaning

All 199 participants' paper maps were individually scanned at 300 DPI. For each paper map, all drawn trajectories were digitized using QGIS 3.20 into an ESRI Shapefile. The following fields were also recorded into an attribute table for each digitized trajectory:

the total *length* (in kilometers) of each recorded trajectory and the *categorization of that trajectory's* purpose as daily/usual or leisure.

Next, using Python (version 3.8.17) and the `geopandas` and `pandas` packages, two analytical trajectory datasets were created for each village, one that contains only leisure trajectories and another that contains daily trajectories. Because each individual's demographic responses are recorded in a dataset separate from their geospatial data, demographic variables also needed to be merged with the trajectory data. Instead of combining all demographic variables with an individual's trajectory information, I used the results from Chapter 5 to help me identify salient demographic features. Based on the results presented in Chapter 5, the salient demographic variables that had the biggest impact on activity differences was, *gender* and *occupation*. Along with the participant's *age*, these were the variables I chose to merge with each participant's trajectory information. This resulted in four analytical datasets with each row containing the coordinates of a single trajectory, the trajectory owner's gender, age, and occupation, and the trajectory's length and purpose categorization.

6.2.3 Using DTW to cluster twelve categories of trajectories

The four analytical trajectory datasets were imported into `R` (version 4.2.1) and processed into a form that could be analyzed by the `dtwclust` package [115]. Using the `tsclust(...)` function, four sets of trajectories were clustered into two to ten clusters. These sets are as follows: 1) Village 1, daily trajectories, 2) Village 1, leisure trajectories, 3) Village 2, daily trajectories, 4) Village 2, leisure trajectories. Next, I disaggregated the trajectories by gender, and clustered the following twelve trajectory sets into two to ten clusters: 1) Village 1 women, daily trajectories, 2) Village 1 men, daily trajectories, 3) Village 2 women, daily trajectories, 4) Village 2 men, daily trajectories, 5) Village 1 women, leisure trajectories, 6) Village 1 men, leisure trajectories, 7) Village 2 women, leisure trajectories, 8) Village 2 men, leisure trajectories. A summary of these clustering activities are shown in Table 6.1. Please note that I chose not to disaggregate the trajectories by occupation, because this resulted in low membership counts within the subcategories (less than 10 trajectories).

Table 6.1: Summary of trajectories clustered for each of the clustering models

Village	**Gender Disaggregation**	**Usage Disaggregation**
Village 1	None	Daily
Village 1	None	Leisure
Village 2	None	Daily
Village 2	None	Leisure
Village 1	Male	Daily
Village 1	Female	Daily
Village 1	Male	Leisure
Village 1	Female	Leisure
Village 2	Male	Daily
Village 2	Female	Daily
Village 2	Male	Leisure
Village 2	Female	Leisure

When implementing `tsclust(...)`, the specific parameter definitions I used are described in Table 6.2. Please note that the package documentation for `dtwclust` describes in detail the mathematical and theoretical foundation behind the various options for model parameterization [108, 115].

Next, the average Silhouette index was calculated for all clusters estimated, for each of the twelve categories listed in Table 6.1. To find the optimal number of clusters for each of the aforementioned categories, the number of clusters with the highest Silhouette index was chosen [108]. The Silhouette index is a measure that assesses how similar an object is to its own cluster, with values ranging from -1 to 1, with higher values indicating better cluster cohesion. A Silhouette index of more than 0.5 indicates a high-quality cluster,

Table 6.2: Parameter definitions used in `tsclust(...)` of the `dtwclust` package

Parameter	Description
type	The choices for this parameter are *hierarchical*, *partitional*, or *fuzzy*. For this application, partitional clustering was chosen to clearly define membership of a given trajectory to a specific cluster.
distance	There are various distance measures that can be specified here. The `dtw_basic` option is the optimized algorithm of the Dynamic Time Warping measure. This was the distance measure chosen for the clustering activity.
centroid	This parameter specifies the prototyping function used to recalibrate and calculate the centroids during the clustering step. Here, the dba option was chosen, representing DTW barycenter averaging. This is an appropriate prototyping function to use when DTW is the chosen similarity measure. The DBA algorithm is described in detail in the documentation [108].

whereas a score of less than 0.5 indicates that the clustering is weak [116]. For example, suppose I wanted to choose between two and eight clusters for a given set of trajectories. The average Silhouette index for the clustering activity with two clusters was 0.5, while eight clusters yielded 0.95. In this hypothetical example, the separation of this specific set of trajectories is better represented by eight clusters than two.

6.2.4 Comparing the radius of gyration and demographic variables among the clusters

Using the `scikit-mobility` Python library I calculated the average radius of gyration (in kilometers) for each of the twelve trajectory sets after the optimal number of clusters was estimated for each [117]. In this analysis, I am using the RG measure to approximate the overall spatial extent of all the defined trajectory sets listed in Table 6.1. This analysis is an addendum to this clustering analysis, and helps contextualize the movement extent of each set of trajectories clustered. Furthermore, the RG measure allows comparisons of overall movement extent between different sets of trajectories.

Descriptive tables were also created summarizing the spatial and demographic features of each cluster. Within each of the twelve categories of trajectories, these tables help describe the demographic and spatial similarities and differences that exist among the estimated clusters.

6.3 Results

These results are presented in four sections. First, a comparison of Silhouette indices is displayed showing the optimum number of clusters for each of these twelve trajectory categories (Table 6.1). Second, after determining the optimal clusters for each clustering category, I show visualizations of the trajectory composition for each. Finally, I compare the radius of gyrations calculated for each of the twelve trajectory categories, to examine differences in mobility extent between village community and gender. Additionally, the demographic composition of each cluster is displayed for each trajectory category.

6.3.1 Using the Silhouette Index to determine the optimum number of clusters for a clustering activity

In Figure 6.1, four subfigures are shown, comparing the Silhouette indices of clustering trajectories based on usage category (daily or leisure trajectories) and village. These initial trajectory sets were not disaggregated by gender. Each trajectory set was clustered nine times, to estimate $k = 2$ to $k = 10$ clusters. This figure shows that for each trajectory category shown, the optimum number of clusters is two for all trajectory categories, except for Village 1 Leisure trajectories, where the optimal number is five clusters. Please note that none of the maximal Silhouette indices for these clustering activities exceeded 0.5, indicating that the clusters found for these trajectory categories are weak.

Next, the trajectory sets are disaggregated by gender, to further examine gender-differences in mobility and movement. Figures 6.2 and 6.3 show eight subfigures comparing Silhouette indices of clustering daily and leisure trajectories from each village, disaggregated by gender. These trajectory sets were also clustered nine times, to estimate $k = 2$ to $k = 10$ clusters. In Village 1, the optimum number of clusters for female/daily, female/leisure, and male/leisure trajectory sets was two clusters, while the optimum number was six clusters for the male/daily trajectory set. In Village 2 the optimum number of clusters for female/daily, female/leisure, male/daily, and male/leisure trajectories was two clusters, for all categories. Here, we found that the strongest clustering structure was found in three trajectory categories, where the maximal Silhouette index exceeded 0.5: Village 1, male, leisure trajectories (SI = 0.65); village 2, female, leisure trajectories (SI = 0.54); and village 2, male, leisure trajectories (SI = 0.52).

6.3.2 Comparisons and visualizations of trajectories within individual clusters

Now that the optimum number of clusters was determined for each trajectory set, the trajectory composition is visualized for each cluster. The next sets of figures show visualizations of aggregated trajectories belonging to each cluster, for each trajectory set. Figure 6.4 shows the daily and leisure trajectories from each village, without gender disaggregation.

The same visualizations are created for trajectory sets disaggregated by gender (Figures 6.5 and 6.6). For some trajectory categories, both Figures 6.4, 6.5, and 6.6 show that the clustering algorithm was able to organize trajectories based on their differential placement within a given village community. This patterning is most evident in the clustering activities with higher Silhouette values. For example, in the subfigures labeled "Village 1: Male, Leisure", "Village 2: Male, Leisure", and "Village 2: Female Daily", clusters are distinctly separated by east/west or north/south halves of their corresponding villages.

6.3.3 Comparison of average radius of gyration and cluster composition

The RG measure was used to quantify an aggregated measure of group mobility. Figure 6.7 shows boxplots of the RG measure calculated for each village, according to usage category (daily vs leisure). Overall, Village 2 has lower RG measures than Village 1 for both daily and leisure trajectories. In Village 1, the median RG for daily trajectories was 0.32km (IQR 0.32-0.33), while in Village 2 the median was 0.26km (IQR 0.25-0.28). For leisure trajectories, the median RG in Village 1 was 0.23km (IQR 0.21-0.35) while in Village 2, the median RG was 0.22km (IQR 0.21-0.23). In Village 1, the RG for leisure trajectories has a much higher extent than daily trajectories. In comparison, in Village 2 the difference in mobility extent is much less pronounced.

When the trajectories are further disaggregated by gender, increasingly granular relationships are revealed between the different trajectory categories, especially between genders. Figure 6.8 shows boxplots of the RG measure calculated for each village, disaggregated by usage category (daily vs leisure) and gender. In both villages, the variation and range of daily RG values are greater for men than women, in both villages. This difference is most distinct in Village 1, where the interquartile range for women's daily activity ranges from 0.310 km to 0.321 km, while for men, the RG varies from 0.242 km to 0.339 km. Similarly, the variation of leisure RG values are greater for men than women, in both villages. For example, in Village 1 the interquartile range for women's leisure activity (median 0.294, IQR 0.275-0.312) is less than men (median 0.328, IQR 0.301-0.354). In general, the RG range for women is less than that of men, at both village sites.

Table 6.3 shows the demographic composition of the clusters from trajectory sets organized by usage category and village site. In each of the clusters shown here, there is a higher number of unique women contributing to the trajectories in each cluster, than men. Tables 6.4 and 6.5 show the demographic composition of the clusters from trajectory sets disaggregated by usage category, village site, and gender. Not many notable differences in cluster characteristics are displayed.

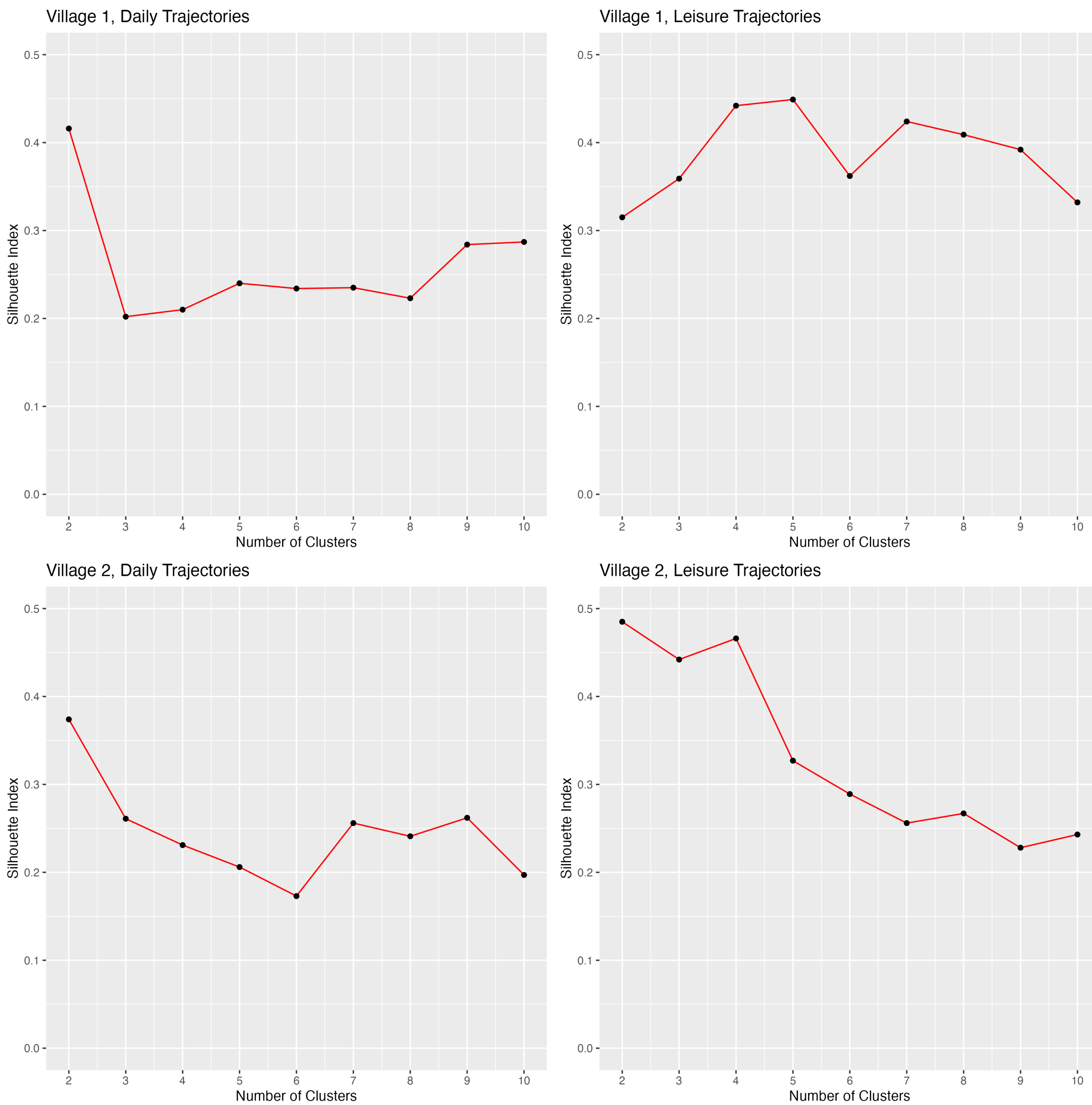


Figure 6.1: Comparison of Silhouette Indices for trajectories organized by village and usage category

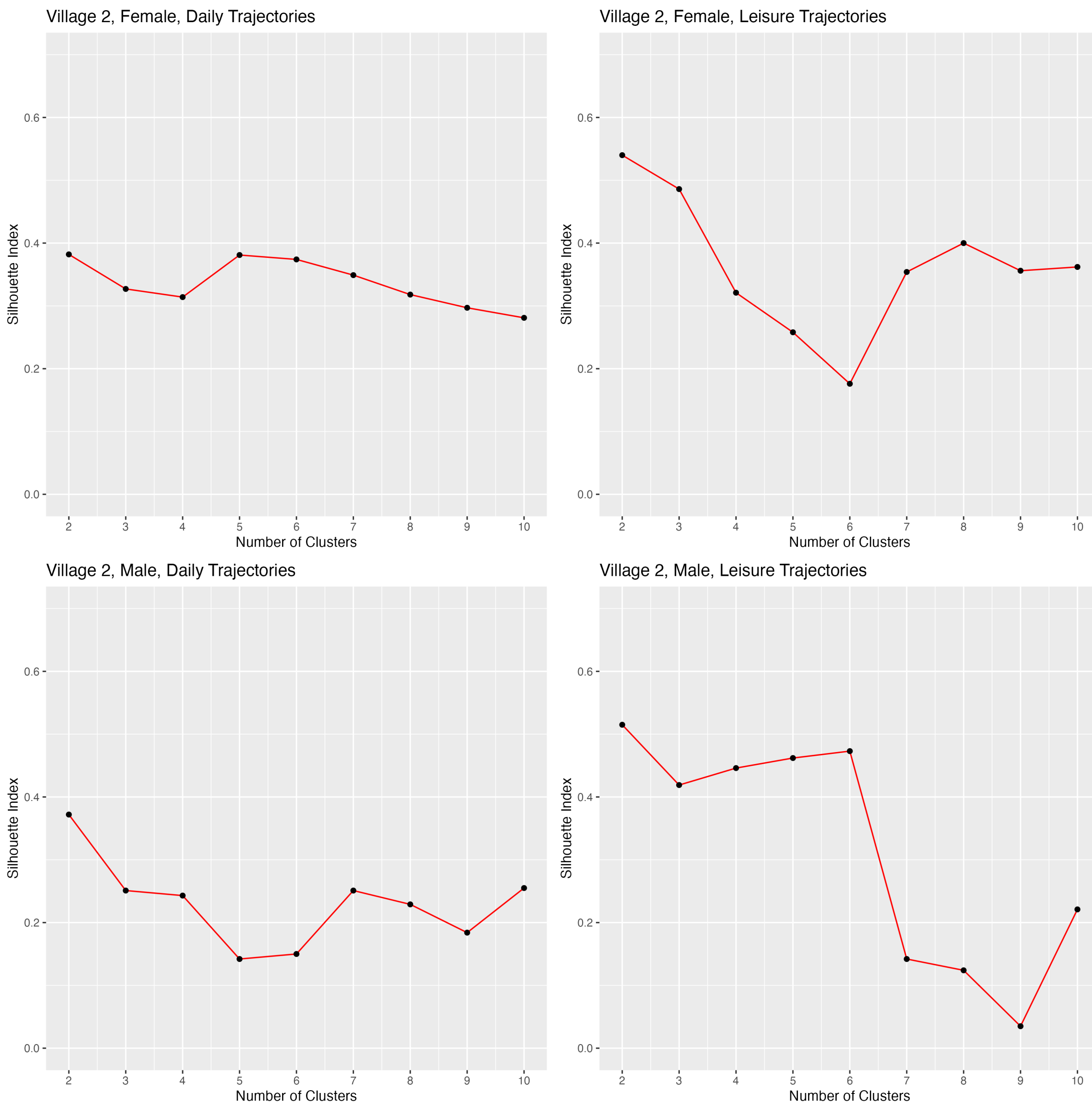


Figure 6.2: Comparison of Silhouette Indices for trajectories organized by usage category and gender – Village 1

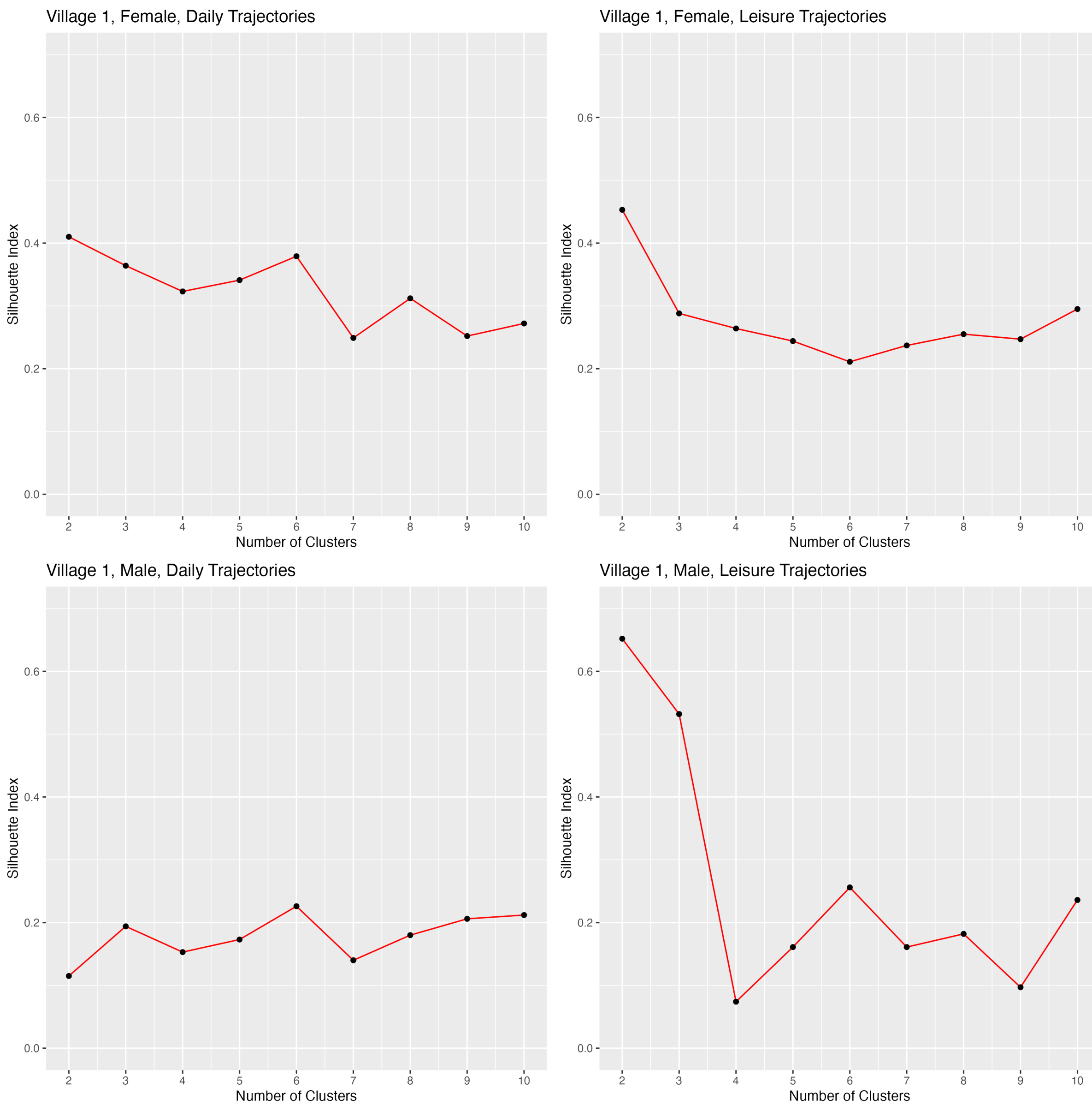


Figure 6.3: Comparison of Silhouette Indices for trajectories organized by usage category and gender – Village 2

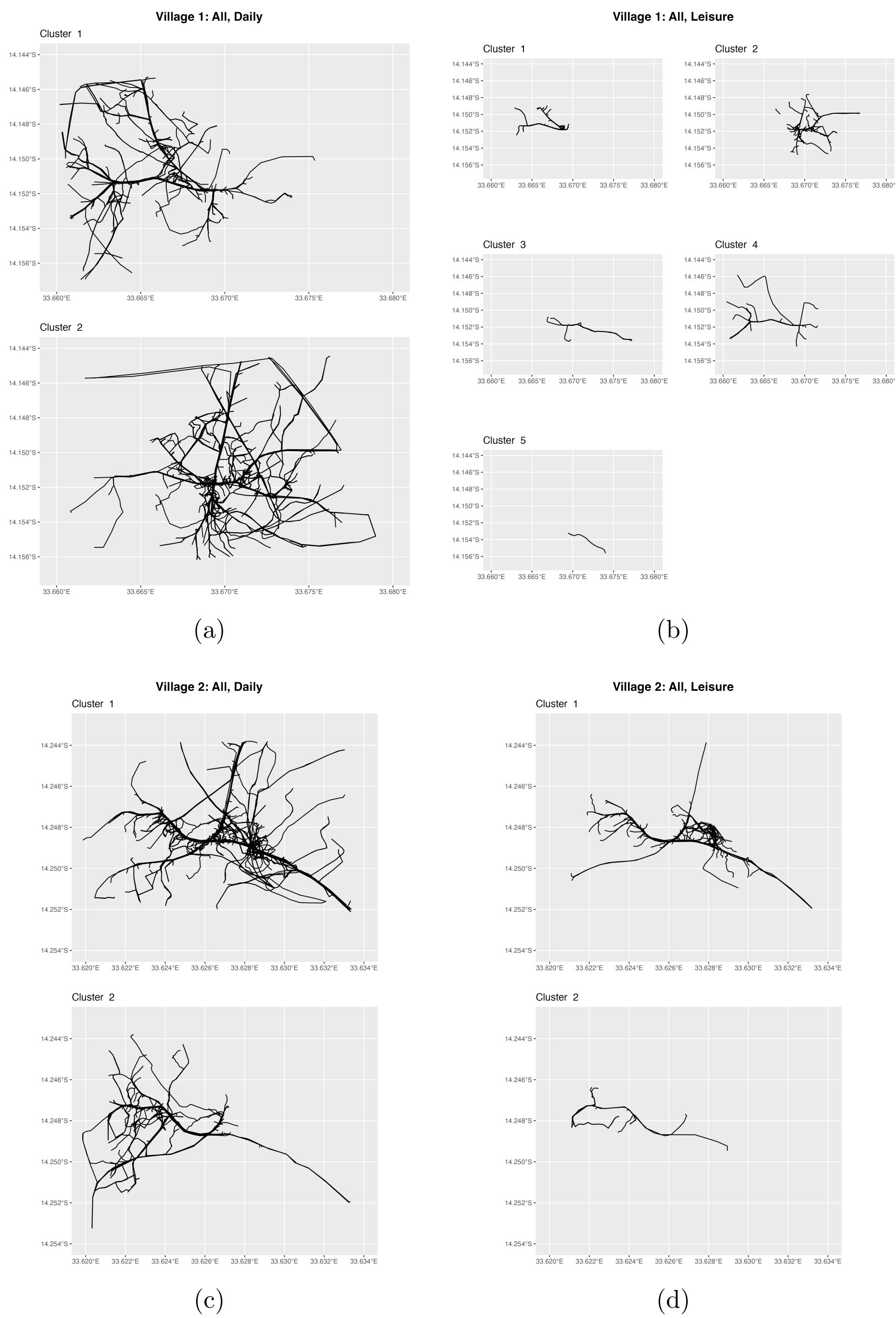
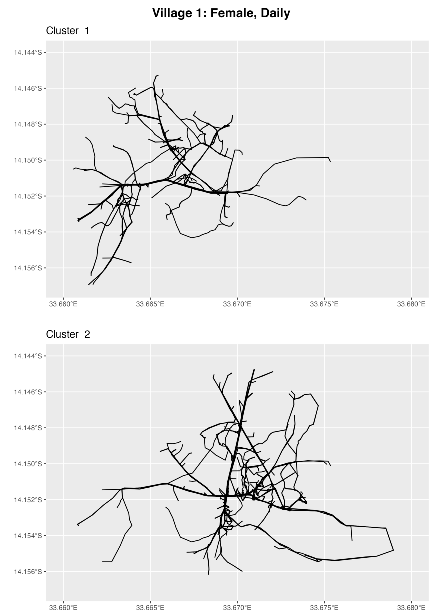
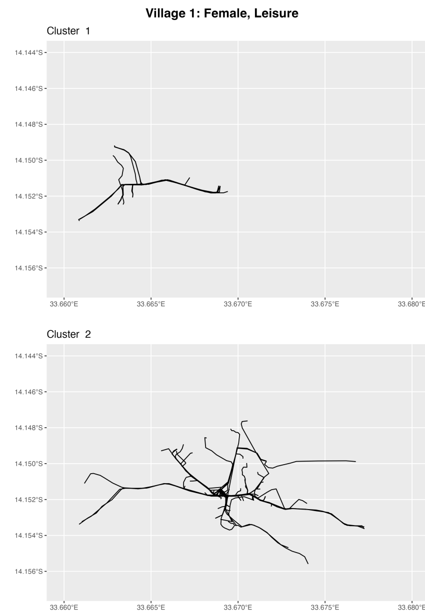


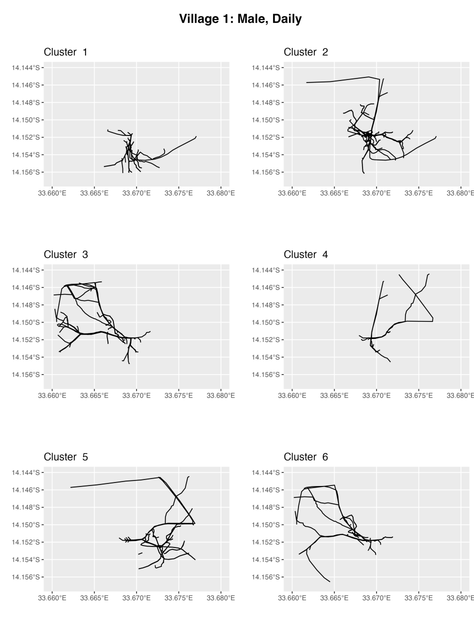
Figure 6.4: Clustered trajectories of trajectory sets organized by village and usage category, without gender disaggregation



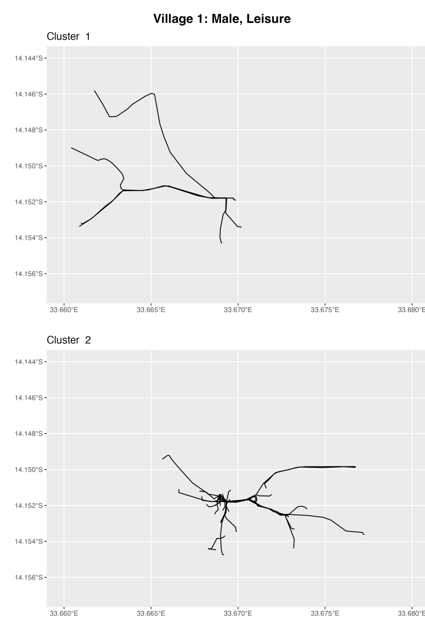
(a)



(b)



(c)



(d)

Figure 6.5: Clustered trajectories in Village 1, disaggregated by usage category and gender

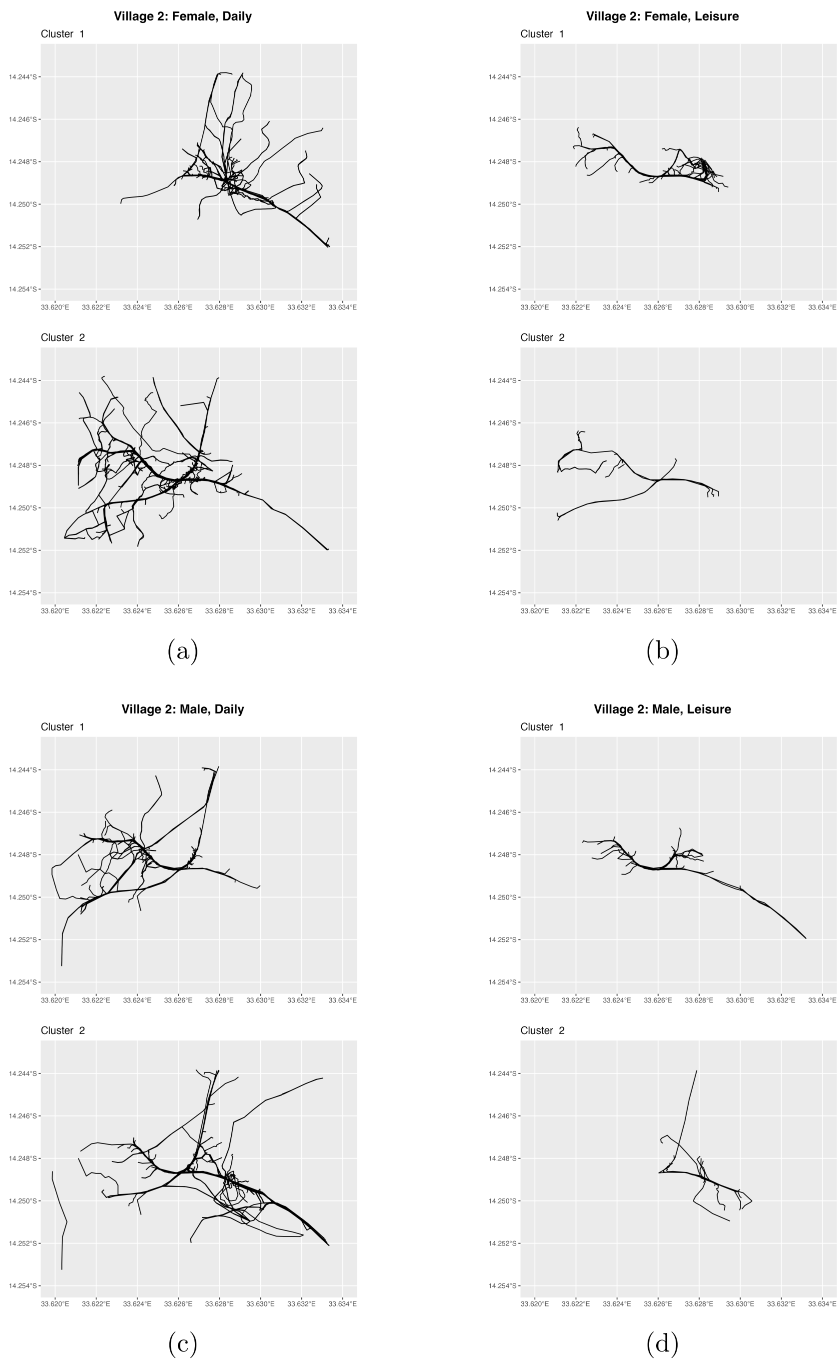


Figure 6.6: Clustered trajectories in Village 2, disaggregated by usage category and gender

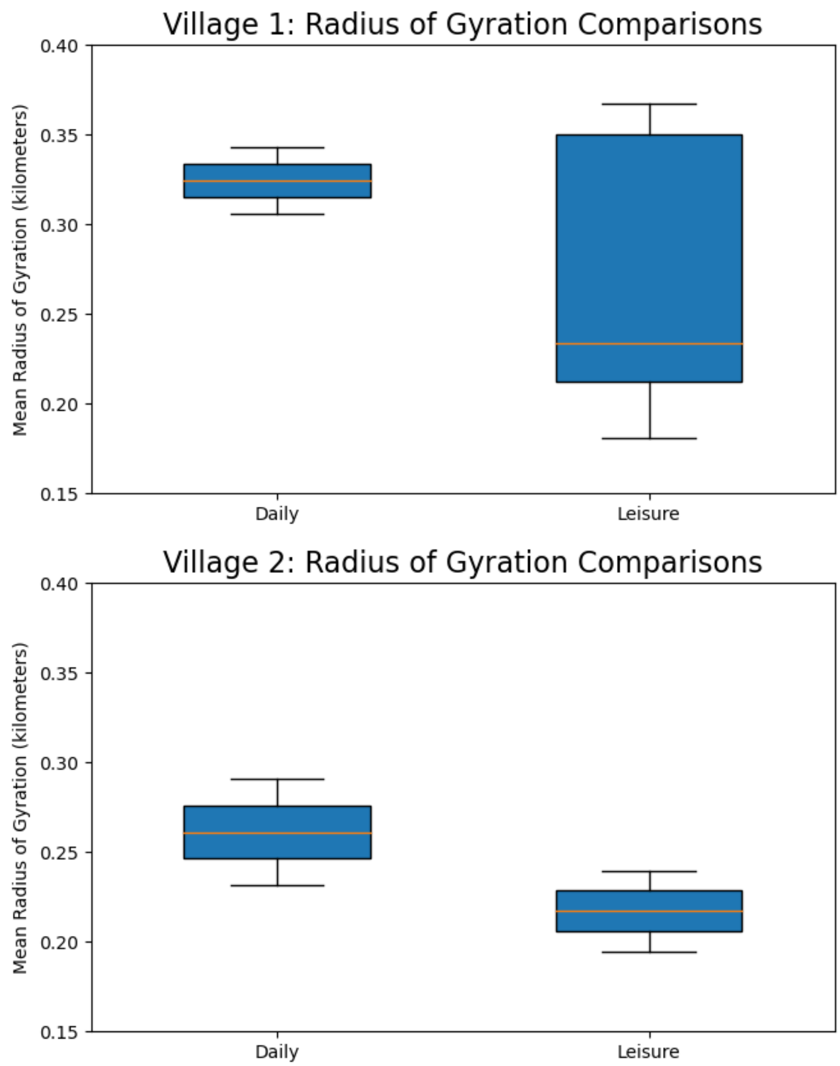


Figure 6.7: Radius of gyration comparisons between usage categories, separated by village site

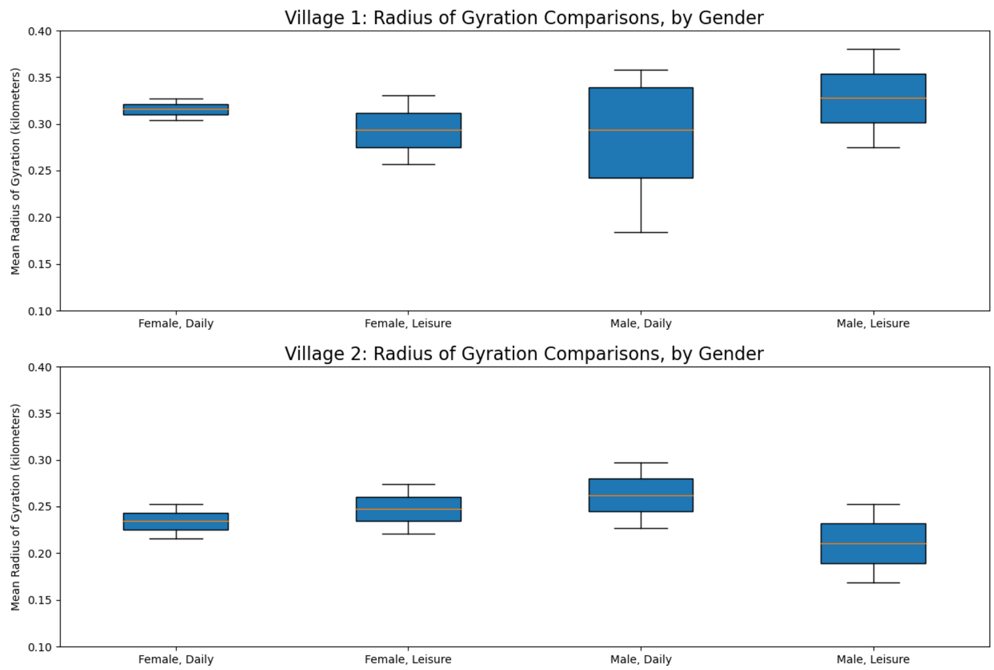


Figure 6.8: Radius of gyration comparisons between usage categories and gender, separated by village site

Table 6.3: Cluster composition for trajectory sets separated by usage category and village site

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Number of men	Number of women	Radius of gyration
1	167	52	0.578	38.4	21	31	0.343
2	365	73	0.423	37.8	32	41	0.306

(a) Village 1, All, Daily

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Number of men	Number of women	Radius of gyration
1	17	14	0.543	45.4	2	12	0.234
2	51	45	0.318	38.0	19	26	0.212
3	4	4	1.140	39.8	1	3	0.350
4	10	9	1.087	33.9	3	6	0.367
5	1	1	0.601	23.0	0	1	0.181

(b) Village 1, All, Leisure

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Number of men	Number of women	Radius of gyration
1	379	77	0.379	39.4	29	48	0.291
2	118	36	0.361	36.5	13	23	0.231

(c) Village 2, All, Daily

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Number of men	Number of women	Radius of gyration
1	55	52	0.482	36.5	21	31	0.240
2	9	8	0.355	36.8	1	7	0.195

(d) Village 2, All, Leisure

Table 6.4: Cluster composition for "daily", organized by gender and village

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Radius of gyration
1	133	31	0.446	37.6	0.327
2	213	41	0.427	33.1	0.304

(a) Village 1, Female, Daily

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Radius of gyration
1	24	10	0.390	43.0	0.184
2	68	27	0.337	43.0	0.231
3	30	21	0.895	40.3	0.358
4	8	7	0.793	48.4	0.309
5	36	14	0.558	48.2	0.278
6	20	13	0.767	41.9	0.349

(b) Village 1, Male, Daily

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Radius of gyration
1	95	23	0.352	38.2	0.216
2	233	40	0.319	39.1	0.253

(c) Village 2, Female, Daily

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Radius of gyration
1	59	21	0.463	41.4	0.227
2	110	27	0.469	38.6	0.297

(d) Village 2, Male, Daily

6.4 Discussion and conclusion

The research question I pose for this chapter is as follows: Are there distinct patterns or clusters of movement among village residents? If so, what are the spatial and demographic characteristics that separate and distinguish these clusters? The answer to this question is "maybe". Given the exploratory nature of applying DTW to a trajectory

Table 6.5: Cluster composition for "leisure", organized by gender and village

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Radius of gyration
1	10	6	0.699	36.7	0.257
2	45	36	0.461	34.0	0.330

(a) Village 1, Female, Leisure

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Radius of gyration
1	4	3	1.283	33.7	0.380
2	24	22	0.358	47.6	0.275

(b) Village 1, Male, Leisure

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Radius of gyration
1	30	29	0.416	35.9	0.221
2	10	9	0.423	35.2	0.274

(c) Village 2, Female, Leisure

Cluster ID	Number of Trajectories	Unique Participants	Trajectory Length (km)	Age (mean)	Radius of gyration
1	16	14	0.542	39.1	0.253
2	8	8	0.525	32.1	0.168

(d) Village 2, Male, Leisure

dataset collected from rural Malawians, it is necessary to first discuss the limitations that may cloud the interpretation of this study's findings. Next, I describe the findings of this study and their implications; and finish a discussion of possible directions for future research.

6.4.1 Limitations

There are two core limitations to this study. The first is the self-reported nature of the collected trajectories and small sample size of the overall trajectory set. I had anticipated that this could be a considerable barrier to collecting reliable and consistent trajectory data. As a result, I attempted to ameliorate inconsistencies by training hired RAs to record movement trajectories themselves, rather than have participants draw and trace their paths. Even with this strategy, there was still variation in precision between the trajectories drawn by the individual RAs.

In the case of one particular male RA, the writing was so unclear that those data could not be digitized. As a result, the male villager trajectory data included in this analysis is considerably fewer than that of women (355 unique trajectories for men versus 674 for women), and the mobility data of 38 male participants could not be included. This may have contributed to the lack of male representation in the first clustering activity, where DTW clustering was applied on trajectories disaggregated by village and use case, but not gender. In these clustering activities, pronounced female membership was observed in some clusters, while male membership was diminished. Had the missing data for male villagers been included, higher male membership may have been observed in certain clusters. The second limitation of this analysis is the application of DTW clustering within this research setting. When I applied DTW clustering separately on each village's trajectory set (disaggregated by use-case but not gender), I observed a high proportion of women in all clusters. Although the lack of pronounced male representation could be due to sample size limitations (discussed earlier), another possible explanation could lie within the mechanics of the DTW algorithm itself. Given that the findings from Chapter 5 demonstrated activity-based differences among gender, I was expecting the resulting mobility clusters to present more variation in gender composition.

The DTW algorithm captures morphological similarity between a pair of trajectories. DTW clustering was able to capture location-specific patterns of movement within the village communities. Similarly shaped and placed trajectories are grouped together and usually clustered around major roads or common areas of a village community (Figures 6.2 and 6.3). However, for the DTW clustering to generate clusters with distinct demographic differences, those different categories of individuals (i.e. men vs women) would need to travel on similar paths, in similar ways, and to similar locations. Although men and women were found to have different activity patterns, this difference does not immediately translate to distinct differences in mobility or ways of moving. For example, a male villager could travel along a major arterial road to reach a plot of agricultural land for his farming work. A female villager could travel along this same road to visit a clinic for her child. Both of these travel trajectories could be grouped in the same cluster, since they are morphologically very similar. Because the DTW clustering algorithm groups trajectories based on morphological similarity alone, the context and motivation behind that movement is not captured. What the DTW clustering misses is the motivations and impetus behind these villager's travel trajectories, which may be important considerations for health and policy makers.

Third, there are limitations to the RG measure within this specific analytical context. Recall that that the RG measure considers all points visited by a given person over a defined period of time. Using these points, a *center of mass* measure is calculated to reflect the weighted mean point of all of the locations visited by the individual. Next, the distances of each location to their center of mass is calculated and then averaged to generate the RG.

For the RG to accurately reflect the spatial extent of an individual, these points would ideally be collected on regular and frequent intervals. Although a high-accuracy GPS device could achieve this task, the methods used in this protocol do not have the capability to collect regular location information. Suppose a participant walks from their home to the village market, and spends two hours there. A GPS device could capture that this individual walked steadily along a path to the village by showing regularly spaced points along the path to the market, and spent a hours at the market, given the density of points that would be collected at the market. However, because the trajectories in my study were drawn and digitized by hand, the digitized points would show the path to the market, but would miss the density of points at the market. Therefore, the RG measure calculated between these two data collection methods would differ considerably: the RG value would be *smaller* for the points collected by GPS device, reflecting that the center of mass value would be closer to the tight collection of points at the market. For the points digitized by hand, the RG value would be larger, because center of mass value would be somewhere along the middle of the path to the market, and would *not* reflect the density of points at the marketplace. This is an important consideration when examining the RG values calculated for this data, that the irregularities that inevitably come with hand-ditizing movement data can reflect artificially large RG values for a given trajectory set.

Fourth, these clustering analyses did not exclude outlier trajectories, given sample size limitations. Potential outliers primarily include trajectories that lead to locations off of the printed village map. For these trajectories, the additional distance was approximated by digitizing lines that wrapped the extent of the paper map boundaries. These additional off-site trajectories would be approximated in length, to estimate the distance to the offsite location from the moment it leaves the paper map extent. The clustering analysis that used these trajectories could be biased by these approximations, especially if the off-paper locations are in drastically different places.

Chapter 7

Conclusions

7.1 Introduction

This dissertation first proposed a methodology to collect community and individual-level geospatial data within rural villages in central Malawi. Then, I presented three analytical tools that parsed these data for spatial insights. At the time of writing, this dissertation presented the first geospatial methodology and clustering analyses to be applied within a Global South context. What, then, are the lessons learned, implications, and potential impacts of this work?

7.2 Overall conclusions

The core theoretical framework for this dissertation is the geographical component of the social determinants of health framework (SDoH). The role of geography in public health has traditionally focused on neighborhood effects on health. However, the neighborhood effects model assumes people spend most of their time in and around their home, and that the majority of health risk and exposure comes from the person's neighborhood.

However, the reality is that most individuals spend the majority of their day outside and away from their home, visiting a range and variety of locations depending on their role within their community, employment requirements, or familial obligations. Heterogeneity of individual mobility and activity exists within any community. This was certainly observed within the two village sites studied in this dissertation. The daily activity of the village residents shows that their day-to-day environmental exposures may be different, given that each resident had unique activity profiles. Further, by tracking the residents' individual movement, I was able to visually observe where my participants visited on a regular basis, and the path they used to reach those locations.

When examining individual health and exposure within the geographical component of SDoH, the *activity space* concept describes the locations and spaces an person interacts with as a result of their daily activities. It is a dynamic measure that has the potential to strengthen and validate the causal relationship between the environmental exposure and disease outcomes, if the methods and research questions are thoughtfully conceptualized. Within this theoretical framework, this dissertation aimed to hone in on three distinct facets of a person's individual geography, 1) their overall cultural and situational *context* and lived environment, 2) their regular *activity* and tasks, and 3) their day-to-day *movement* or spatial extent.

The biggest strength of this thesis is the way it describes the complete *activity space* of the recruited participants. This is achieved with the comprehensive focus group discussions outlining community movement, culture, and history in Chapter 4, an examination of activity archetypes in Chapter 5, and the analysis of community movement in Chapter 6. These three analyses help present the totality of individual geography for Malawian villagers, and individually address each of these geographical facets in a rural Malawian setting. In Chapter 4, I used a guided focus group questionnaire to collect community-level cultural, demographic, and seasonal movement and activity insights. This activity elucidated community demographics, cultural practices, seasonal and daily schedules that were relevant to individual activity and mobility. In particular, this activity clearly demonstrated gender differences in daily tasks, occupation, and chores.

In Chapters 5 and 6, I honed in on the individual villager: Chapter 5's focus was on individual activity and Chapter 6 focused on individual movement. In Chapter 5, I used latent profile analysis (LPA) to extract latent patterns of activity, from three different activity labeling groupings. This analysis demonstrated the importance of context-dependent activity labeling, because each of the three labeling groupings revealed their own individual latent categories and profiles. From this analysis, I found that the culturally-relevant labeling grouping was able to produce the most distinct activity profiles that I named "Chillers", "Doers", and "Balanced". "Chillers" were mostly men, "Doers" were mostly women, and "Balanced" had an equal proportion of each gender. The culturally-relevant labeling grouping was able to abstract the distinct gender roles that was already observed from the focus group discussions described in Chapter 4.

Chapter 6's focus was individual movement trajectories, and exploring the usage of dynamic time warping as a similarity measure, to cluster paths and tracks collected from this study's participants. This analysis demonstrated that DTW clustering algorithm groups trajectories on morphological similarity, and may be able to capture location-specific patterns of movement within the study villages. Similarly shaped and located trajectories and grouped together, usually clustered around major roads or common areas of a village community.

7.3 Research contributions towards health geography and public health

The lessons learned from this dissertation may be impactful for HIV implementation policy. The results from the analyses in Chapters 5 and 6 suggest that gender-based interventions could be an effective strategy to expand access to HIV testing and treatment in rural Malawi. This suggestion is consistent with a general trend that men are less likely to enter the HIV testing and treatment cascade across sub-Saharan Africa [118–120]. A study that characterized successful strategies to encourage testing among men showed that place-based outreach efforts, including door-to-door testing services and workplace testing outreach, have seen success in engaging men in HIV testing and treatment. Therefore, the *access point* of these outreach services could be a core determinant in the success of an HIV testing and treatment program. In this analysis, I found that men and women often occupy different leisure spaces, where men are more likely to convene in bars and play board games together while women are often found gathering with friends at home. Compared to women, men are also more likely to work in the transportation sector as drivers or bicycle taxis. These differences in spatial context could help health providers reach the remaining untested PLHIV, in countries and communities that are overperforming on their 95-95-95 testing and treatment targets.

To address the second component of the aforementioned question, this dissertation addressed a core gap within the breadth of literature describing geography and SDoH. As described in Chapter 2, rural Global South communities are understudied within the field of public health research, even though a variety of economic, transportation, and health measures show that rural communities often lack access to health and social services. Furthermore, even when rural communities are targeted for public health research, the study protocols are often led by western researchers who impose their own biases and views on these communities. This dissertation aimed to capture the most complete picture of activity and mobility within the villages sampled for this research, by capturing the aforementioned three facets of individual geography. Future movement and activity research should explore similar approaches for refining and developing geospatial surveys for their specific study area. Given the uniqueness of individual and community-level activity and mobility between villages, regions, and countries, such an approach can reveal the wide heterogeneity of individual human geography, and the relationship of this geographic heterogeneity with various health exposures and risks.

Appendix A

Phase 1 Focus Group Questionnaire (English Version)

PHASE 1: *Field Checklist and Role Assignments*

Document Summary

This field checklist contains key tasks that need to be completed prior to beginning the FGDs; and essential deliverables expected from the FGDs after data collection is complete..

Assignment of FGD Roles

Male Team

Role	**Name of Research Assistant**	**Assigned Mobile Number**
Moderator		
Note Taker 1		
Note Taker 2		

Female Team

Role	**Name of Research Assistant**	**Assigned Mobile Number**
Moderator		
Note Taker 1		
Note Taker 2		

1. Before Departure to Field Site

Each field team has all necessary ***paper documents***, with extra copies:

- ☐ a packet of flipchart paper for **participatory mapping** and **brainstorm** activities
  - at least **four sheets** are needed per team
- ☐ 3 copies of the ***Response Booklet - Note Taker Version***
  - **two** for real-time data collection
  - **1** for consolidation
- ☐ 3 copies of the data collection checklist

Each field team has all the necessary ***refreshments for the participants***

- ☐ 10-15 total Fanta/Coke refreshments per focus group
- ☐ 10-15 total bottled water per focus group
- ☐ 10-15 biscuit packs per focus group

Each field team has all necessary ***data collection supplies and stationary***:

- ☐ 10-15 copies of the informed consent language
- ☐ extra ream of blank paper (for participants to jot down notes)
- ☐ colored markers
- ☐ 2 notebooks - one for each of the notetakers (2 RAs per focus group)
- ☐ 5-6 pens

2. Before Beginning the FGD

- ☐ Find and prepare the FGD venue.
- ☐ Set out documents and prepare the FGD space.
- ☐ Make sure participants are comfortable.
  - Set out mats, pillows, or blankets, if needed. This is especially important for the female focus groups, since some women may have young children.
- ☐ Distribute participant ID signs for all participants using printer paper.
- ☐ Deliver the consent language, and fill out the ***Phase 1 Research Participation Agreement***, page 4 of the ***Consent to Participate in WeMap FGD*** document.

3. Right After the FGD

- ☐ Immediately debrief with your field team to consolidate FGD responses.
- ☐ Take a ***photo with good lighting*** of the following documents and send them to the "WeMap Phase 1 Field Team" *WhatsApp* group.
  - the community map
  - ***Phase 1 Research Participation Agreement***
  - all pages of the consolidated ***Response Booklet - Note Taker Version***
- ☐ Compose a voice memo with updates on how the day's data collection went. Send this voice memo to the "WeMap Phase 1 Field Team" *WhatsApp* group.
- ☐ Leave both male and female community maps with the community.

PHASE 1: Response Booklet - Note Taker Version

Version (check one)

☐ *Note Taker Name:* ___________________________

☐ *Consolidated Version*

Participant Roster

Participant ID	**Age** (Years)	**Marital Status** (Married / Single)	**Main Source of Income**	**Years Lived in Village**
Participant 1		Married ☐ Single ☐		
Participant 2		Married ☐ Single ☐		
Participant 3		Married ☐ Single ☐		
Participant 4		Married ☐ Single ☐		
Participant 5		Married ☐ Single ☐		
Participant 6		Married ☐ Single ☐		
Participant 7		Married ☐ Single ☐		
Participant 8		Married ☐ Single ☐		
Participant 9		Married ☐ Single ☐		
Participant 10		Married ☐ Single ☐		
Participant 11		Married ☐ Single ☐		
Participant 12		Married ☐ Single ☐		

I. Preparation

II. Community Mapping Initiation

1.	Prior to starting the FGD, first have people lay out the main geographical and structural features of the village: ***Mapping*** ☐ First have the participants ***situate the venue*** for the FGD interview within the larger context of the ***community boundaries***.

2.	***Mapping*** ☐ Now, let's set the legend for a number of key features using a **blue marker**. Key features could be places that everyone knows and/or everyone goes to or visits frequently.

III. Opening and Background

1.	Tell me about the history of this community. ● When was it established? ● Who were the founding members or tribes? ● What are the existing tribes? ● How many languages are spoken in this community? ● How many religions are represented within the village? ● Was/is poloygamy common?

When was it established?

Date ______________________

Who were the founding members or tribes?

Founding members 1. ______________________

2. ______________________

3. ______________________

4. ______________________

Founding tribes 1. ______________________

2. ______________________

3. __________
4. __________

What are the existing tribes?

Existing tribes

1. __________
2. __________
3. __________
4. __________

How many languages are spoken in this community?

Number of languages __________

List languages

1. __________
2. __________
3. __________

How many religions are represented within the village?

Number of religions __________

List religions

1. __________
2. __________
3. __________
4. __________

Was/is poloygamy common?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Comments

2.	Describe the typical family in this village. • Do all members of a family reside in the same home? • Is it common for certain members of the family to sleep in one home certain nights, and other homes other nights?

Do all members of a family reside in the same home?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Comments

Is it common for certain members of the family to sleep in one home certain nights, and other homes other nights?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Comments

3.	Think about the walkways, roads, and paths within your village: ***Mapping*** ☐ On the map, please draw the major footpaths with a **dotted black line** (**- - -**), and vehicle roadways with a **solid black line** (_). ***Questions*** ● How were these paths made or constructed? Or did they just appear? ● Are there paths that are regularly maintained by work parties? ● Do paths change with respect to season? In other words, come rainy season, do some paths disappear and some new ones appear?

How were these paths made or constructed? Or did they just appear?

Indicate **One** *Option*

Mostly constructed ☐
Mostly appeared ☐
A mix of both ☐

Comments

Are there paths that are regularly maintained by work parties?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Comments

Do paths change with respect to season? In other words, come rainy season, do some paths disappear and some new ones appear?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Comments

IV. Social Structure and Village Schedules

1.	• Are there different groups within your community, or individuals who interact more frequently with each other? • Are these groups religious, tribal, occupational, recreational, or gender-based? • *If none of the above, how would you describe the common characteristics that describe these groupings?*

WeMap | **Phase 1 Response Form | Note Taker Version, ENG | Date:** February 9, 2020

Are there different groups within your community, or individuals who interact more frequently with each other?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Comments

Are these groups religious, tribal, occupational, recreational, or gender-based?

Indicate **All** *Options*

Religious ☐
Tribal ☐
Occupational ☐
Recreational ☐
Gender-based ☐

Comments

If none of the above, how would you describe the common characteristics that describe these groupings?

Comments

9

2.	Think of the different groups that you identified in the previous question: ***Mapping*** ☐ On the map, please indicate the parts of the village where different social groups reside (e.g. tribal and religious) with a **red marker**. ***Questions*** ● Do people belonging to different groups tend to live in a certain area of the village?

Do people belonging to different groups tend to live in a certain area of the village?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

List groupings

1. ____________________
2. ____________________
3. ____________________
4. ____________________

Comments

***WeMap* | Phase 1 Response Form | Note Taker Version, ENG | Date:** February 9, 2020

3.	Please describe the village scheduling for a normal week. Think of the week prior to this interview: ● Is there a market day for the village? If yes, when is market day? ● What days do most people visit church? ● What other events happen regularly during the week?

Is there a market day for the village? If yes, when is market day?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Market day ____________________

What days do most people visit church?

Church days
1. ____________________
2. ____________________
3. ____________________
4. ____________________

Comments

What other events happen regularly during the week?

Other weekly events
1. ____________________

11

2. ____________

3. ____________

4. ____________

4.	• Does this weekly schedule change in the dry or harvest/planting season? • If yes, how would the weekly schedule change during the dry season? • During harvest/planting season? • Are there other circumstances where the weekly schedule would change?

Does this weekly schedule change in the dry or harvest/planting season?

Indicate **Yes** *or* **No for dry season** ***Yes*** ☐ ***No*** ☐

Indicate **Yes** *or* **No for harvest season** ***Yes*** ☐ ***No*** ☐

If yes, how would the weekly schedule change during the dry season?

Comments

During harvest/planting season?

Comments

Are there other circumstances where the weekly schedule would change?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Comments

5.	On an average ***weekday*** *(if yesterday was a weekday, ask the focus group to think about yesterday)*, how many people… • ...of a different <u>religion</u> than you, do you talk to? • ...from a different <u>tribe</u> than you, do you talk to?

...of a different <u>religion</u> than you, do you talk to?

Participant ID	*Response*
Participant 1	______________________
Participant 2	______________________
Participant 3	______________________
Participant 4	______________________
Participant 5	______________________
Participant 6	______________________
	______________________

Participant 7 ____________________

Participant 8 ____________________

Participant 9 ____________________

Participant 10 ____________________

Participant 11 ____________________

Participant 12 ____________________

...from a different tribe than you, do you talk to?

Participant ID	*Response*
Participant 1	____________________
Participant 2	____________________
Participant 3	____________________
Participant 4	____________________
Participant 5	____________________
Participant 6	____________________
Participant 7	____________________
Participant 8	____________________
Participant 9	____________________
Participant 10	____________________
Participant 11	____________________
Participant 12	____________________

6.	On an average day during the ***weekend*** *(ask the focus group to think about the past Sunday or Saturday)*, how many people… • ...of a different religion than you, do you talk to? • ...from a different tribe than you, do you talk to?

...of a different religion than you, do you talk to?

Participant ID *Response*

Participant 1 ______________________

Participant 2 ______________________

Participant 3 ______________________

Participant 4 ______________________

Participant 5 ______________________

Participant 6 ______________________

Participant 7 ______________________

Participant 8 ______________________

Participant 9 ______________________

Participant 10 ______________________

Participant 11 ______________________

Participant 12 ______________________

...from a different tribe than you, do you talk to?

Participant ID *Response*

Participant 1 ______________________

____

Participant 2 ______________________

Participant 3 ______________________

Participant 4 ______________________

Participant 5 ______________________

Participant 6 ______________________

Participant 7 ______________________

Participant 8 ______________________

Participant 9 ______________________

Participant 10 ______________________

Participant 11 ______________________

Participant 12 ______________________

7.	• Are there community-wide events that happen on a weekly basis (e.g. market days, church days, etc.)? • Who goes to or participates in these events? • On what days during the week do these events occur?

Are there community-wide events that happen on a weekly basis (e.g. market days, church days, etc.)?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

List events

1. ______________________
2. ______________________
3. ______________________
4. ______________________

5.

Who goes to or participates in these events?

Events	*Participants*
1.	
2.	
3.	
4.	
5.	

On what days during the week do these events occur?

Events	*Happens when?*
1.	
2.	
3.	
4.	
5.	

8.	• Are there community-wide events that happen on a monthly or yearly basis? • Who goes to or participates in these events? • How frequently do these events occur?

Are there community-wide events that happen on a monthly or yearly basis?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

List events 1.

2.

3.

4.

5.

Who goes to or participates in these events?

Events	*Participants*
1.	
2.	
3.	
4.	
5.	

How frequently do these events occur?

Events	*Happens when?*
1.	
2.	
3.	
4.	
5.	

V. Occupational Structure

1.	How do people earn a living in your community throughout the year? • Begin in January and describe the top two monthly occupations for each month until December. Please list occupations for both adult men and women.

Begin in January and describe the top two monthly occupations for each month until December. Please list occupations for both adult men and women.

Month	*Top two occupations - women*	*Top two occupations - men*
January	1. 2.	1. 2.
February	1. 2.	1. 2.
March	1. 2.	1. 2.
April	1. 2.	1. 2.
May	1. 2.	1. 2.
June	1. 2.	1. 2.
July	1. 2.	1. 2.
August	1. 2.	1. 2.
September	1. 2.	1. 2.

October	1. 2.	1. 2.
November	1. 2.	1. 2.
December	1. 2.	1. 2.

2.	*From the list generated in the prior question, flatten the responses to form a list of unique occupations. For each unique job/occupation that was named, ask…* ● where the site of that work is ● how long it takes to travel there from the participant's home ● the means of travel to the work site

Occupation	*Site of work*	*Time traveled*		*Means of travel*	
		1:	2:	1:	2:
		3:	4:	3:	4:
		5:	6:	5:	6:
		7:	8:	7:	8:
		9:	10:	9:	10:
		1:	2:	1:	2:
		3:	4:	3:	4:
		5:	6:	5:	6:
		7:	8:	7:	8:
		9:	10:	9:	10:
		1:	2:	1:	2:
		3:	4:	3:	4:
		5:	6:	5:	6:

		7:	8:	7:	8:
		9:	10:	9:	10:
		1:	2:	1:	2:
		3:	4:	3:	4:
		5:	6:	5:	6:
		7:	8:	7:	8:
		9:	10:	9:	10:
		1:	2:	1:	2:
		3:	4:	3:	4:
		5:	6:	5:	6:
		7:	8:	7:	8:
		9:	10:	9:	10:
		1:	2:	1:	2:
		3:	4:	3:	4:
		5:	6:	5:	6:
		7:	8:	7:	8:
		9:	10:	9:	10:
		1:	2:	1:	2:
		3:	4:	3:	4:
		5:	6:	5:	6:
		7:	8:	7:	8:
		9:	10:	9:	10:
		1:	2:	1:	2:
		3:	4:	3:	4:
		5:	6:	5:	6:

		7: 8: 9: 10:	7: 8: 9: 10:

WeMap | **Phase 1 Response Form | Note Taker Version, ENG | Date:** February 9, 2020

3.	*For each unique job/occupation that was named, also ask…* • ...if men and women both participate in this occupational activity • …[if the answer is "yes"] how do their roles and work roles differ?

Occupation	*Who participates?*	*Role of women*	*Role of men*
	Women ☐ *Men* ☐		
	Women ☐ *Men* ☐		
	Women ☐ *Men* ☐		
	Women ☐ *Men* ☐		
	Women ☐ *Men* ☐		
	Women ☐ *Men* ☐		
	Women ☐ *Men* ☐		
	Women ☐ *Men* ☐		

4.	*Don't forget to probe about concurrent occupations.* • *For example, if the dominant occupation is bricklaying, then does farming of a secondary crop still happen?*

Occupation	*Concurrent Occupations*	*During when or what months?*
	1: 2: 3:	1: 2: 3:

23

	1: 2: 3:	1: 2: 3:
	1: 2: 3:	1: 2: 3:
	1: 2: 3:	1: 2: 3:
	1: 2: 3:	1: 2: 3:
	1: 2: 3:	1: 2: 3:
	1: 2: 3:	1: 2: 3:
	1: 2: 3:	1: 2: 3:

VI. Rest Break (insert when needed)

VII. Modes of Transportation

WeMap | **Phase 1 Response Form** | **Note Taker Version, ENG** | **Date:** February 9, 2020

1.	Together as a group, please come up with a list of locations you visit regularly. By *regular*, we mean on a daily, weekly, or monthly basis.

Regularly visited locations
1. __________
2. __________
3. __________
4. __________
5. __________
6. __________
7. __________
8. __________
9. __________
10. __________

2. ***Questions***

- Now, for all these locations you indicated, how frequently do you visit them (e.g. daily, weekly/bi-weekly, or monthly)?

Mapping

- ☐ To distinguish between the **blue marked** locations indicated in section II, please use a **green marker** for these locations.
- ☐ Please distinguish between daily, weekly, and monthly locations, by using a triangle to mark daily locations, squares for weekly, and circles for monthly.
- ☐ Please take note of locations that are *within* and *outside* the village.

Now, for all these locations you indicated, how frequently do you visit them (e.g. daily, weekly/bi-weekly, or monthly)?

Location	Frequency of visit
1.	

25

2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	

3.	Please indicate for every location, how you travel (e.g. on foot, bicycle, motorcycle taxi) and why you travel to this place.

Location	*Mode of transport*	*Why visit?*
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		

9.		
10.		

VIII. Gender and Age Roles

1.	Think about a typical day for a villager in your village. • How does this day differ between men and women?

How does this day differ between men and women?

Typical day for an adult man
Typical day for an adult woman (**without** children)
Typical day for an adult woman (**with** children)

2.	Think about your friends and acquaintances in your village. ● Are there adult individuals you know who want to participate in chores and work of daily life but are unable to because of their age? ● Do these individuals have a common age? About what age do people start being unable to participate in chores and work? ● Do these individuals have a common medical condition? What is it? ● If these individuals do not participate in the chores and work of daily life, what do they do during the day? What are the locations that they frequently visit? How would they get to these places?

Are there adult individuals you know who want to participate in chores and work of daily life but are unable to because of their age?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Comments

Do these individuals have a common age? About what age do people start being unable to participate in chores and work?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Common age groupings

1. ____________________
2. ____________________
3. ____________________
4. ____________________

Comments

Do these individuals have a common medical condition? What is it?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Common medical conditions

1. ________________
2. ________________
3. ________________
4. ________________

Comments

If these individuals do not participate in the chores and work of daily life, what do they do during the day? What are the locations that they frequently visit? How would they get to these places?

Daily schedule

Frequented locations	*How are these locations accessed?*
1.	1.
2.	2.
3.	3.
4.	4.

IV. Health and Perceptions of People Living with HIV (PLWH)

1.	• What are some of the major health problems that your community face?

What are some of the major health problems that your community face?

Major health problems	1.
	2.
	3.
	4.

2.	• Where do villagers seek care? • In general, are villagers with these health problems able to seek care from the nearest health facility? If not, where do they go to seek care? • Please come up with a list of health providers. If possible, please include traditional medicine providers.

Where do villagers seek care?

Health care facilities

1. __________
2. __________
3. __________
4. __________

In general, are villagers with these health problems able to seek care from the nearest health facility? If not, where do they go to seek care?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

If not, where do they seek care?

Please come up with a list of health providers. If possible, please include traditional medicine providers.

Health providers	*Traditional medicine / healer? (***Yes** *or* **No***)*
1.	***Yes*** ☐ ***No*** ☐
2.	***Yes*** ☐ ***No*** ☐
3.	***Yes*** ☐ ***No*** ☐
4.	***Yes*** ☐ ***No*** ☐

3.	Let's turn our attention towards a specific health problem, one that affects all people in

	Malawi: HIV. • Do you think HIV is a problem within your community? Why or why not?

Do you think HIV is a problem within your community? Why or why not?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Why or why not?

4.	• Can people living with HIV be open about their status to members of your village? Why/Why not?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Why or why not?

5.	• Are there traditional healers or community health centers located within your community? • Who goes to traditional healers and why? • Who goes to community health centers and why?

Indicate **Yes** *or* **No** ***Yes*** ☐ ***No*** ☐

Who goes to traditional healers and why?

Comments

Who goes to community health centers and why?

Comments

X. Generalization and Comparison with Surrounding Communities/Villages

1.	Think of the villages around yours: ● Can you name 3-4 of these villages?

Can you name 3-4 of these villages?

List villages

1. ______
2. ______
3. ______
4. ______

2.	Think of everything that we've discussed so far, specifically, social structure, occupations, road and path infrastructure.

	• How is your village different or similar to the villages around yours?

How is your village different or similar to the villages around yours?

Comments

Appendix B

Phase 2 Geospatial Survey Instrument (English Version)

PHASE 2: *Survey*
English Version

RA Name: ______________________________

Participant Code: ______________________________

Village Census *(Complete Weekly)*

From the chief, please enquire about the total above 18 village population. Also find out how many adult men and women are in the village.

Village name ________________________

Total population ________________________

Total women ________________________

Total men ________________________

Daily Preamble *(Complete once per day of data collection)*

Please describe any village events that are happening today. This includes, whether or not today is a market day, if there's an ongoing funeral, or another village event.

Describe what's happening in the village today

Starting Landmark ________________________

Main Survey Instrument *(Complete with every participant)*

I. Demographics

1. **[BIN]** Gender

   ☐ M
   ☐ F

2. **[NUM]** What is your current age in years?

   ____

3. **[NUM]** Year of birth

   ____

4. **[SA]** Where were you born?

   > **[RA NOTE]** We are not looking for a location (hospital vs village) of birth, but the PLACE of birth.

   Town or Village or City ________________________

   District ________________________

   Country ________________________

5. **[NUM]** How many people currently live in your household?

   ____

6. **[NUM]** Think of the past year, how many *months* did you live in this community, if not the entire year?

   > **[RA NOTE]** Phrase this question like this: "From the date of the interview, think of the past year."

   ____

7. **[NUM]** How long have you lived in this ***census enumeration village*** community (in *years* and *months*)?

   Years _________

   Months _________

8. **[BIN]** Are you currently married?

   ☐ No
   ☐ Yes

9. **[BIN]** Have you ever attended formal education?

   ☐ No (SKIP to 11)
   ☐ Yes

10. **[CAT]** What was the highest year of school you completed?

   ☐ Primary Education
   ☐ Secondary Education
   ☐ Tertiary Education

11. **[CAT]** Religious affiliation

   ☐ CCAP
   ☐ Catholic
   ☐ Anglican
   ☐ Islam
   ☐ Jehovah's Witness
   ☐ Seventh-day Adventist
   ☐ Other __________________

12. **[NUM]** How many living children do you currently have?

   [RA NOTE] This refers to both biological and non-biological children.

   _________ (IF . = 0, SKIP to 14)

13. **[NUM]** How many children currently live with you?

   [RA NOTE] This refers to both biological and non-biological children.

   ____

14. **[NUM]** What is the age of your youngest child (in *years* and *months*)?

    Years _________

    Months ________

15. **[CAT]** What is your tribe?

    ☐ Chewa
    ☐ Lomwe
    ☐ Sena
    ☐ Mang'anja/Nyanja
    ☐ Ngoni
    ☐ Tumbuka
    ☐ Tonga
    ☐ Yao
    ☐ Other ___________________________

16. **[CAT]** Household assets: Does your household have...

    [RA NOTE] Respondents should select all that apply. Also, make sure to emphasize that all assets should be functional.

    ☐ a metal roof?
    ☐ electricity?
    ☐ a paraffin lamp with no glass?
    ☐ a paraffin lamp with glass?
    ☐ a radio?
    ☐ a television?
    ☐ a cellular phone (*not* smartphone)?
    ☐ a cellular phone (smartphone)?
    ☐ a bed?
    ☐ a sofa set?
    ☐ a table?
    ☐ a refrigerator?
    ☐ a mattress?
    ☐ a chair(s)?
    ☐ cattle?
    ☐ goats?
    ☐ sheep?
    ☐ pigs?
    ☐ donkeys?
    ☐ chickens?
    ☐ other poultry?
    ☐ Other ___________________________

17. **[CAT]** How would you describe your current working status?

☐ Business
☐ Working formally (employed full time)
☐ Working informally (*ganyu*, piece work)
☐ Not working

18. **[CAT]** Please think of the past 12 months, how would you describe your primary occupation?

☐ Student
☐ Unemployed
☐ Informal labor
☐ Subsistence farming
☐ Laundry
☐ Food and vegetable stalls
☐ Bicycle taxi
☐ Other ___________________

19. **[NUM]** Think back on the past month, and think of your current work schedule. Even if you're not working now, think of the time when you were most recently working. On average, how many days do you/did you work on a given week?

_________ days worked per week

20. **[BIN]** Do you currently have a disability that affects your day-to-day mobility?

☐ No (SKIP to 22)
☐ Yes

21. **[CAT]** With your current disability...

[RA NOTE] Read the response options to the respondent.

☐ ...I *can* move independently, *without assistance*
☐ ...I *can* move independently, *with assistance*
☐ ...I *cannot* move independently, even with assistance

II. Travel Behavior Journal

[RA NOTE] *Prior to starting this task, you will need to link the paper map with the respondent's digital responses. On the back of the map map, first write* ***Participant Code****. At the end of the tablet survey, you will be asked to save the survey with this* ***Participant Code****.*

[Script] Now, I'm going to ask some questions about your village community and your daily activity and movement. On a paper map, we will be drawing locations you visit frequently. When going through this exercise, we are defining your village as your ***census enumeration village (Ndumila 1 or Mamina)***. We are asking you to be as precise as possible with your responses, specifically when we start drawing paths between locations on the map. Are you ready?

Movement within Village

22. **[MAP]** I will now ask you about how you define your ***census enumeration village (Ndumila 1 or Mamina)*** on this map. Help the research assistant as they draw a single boundary line around the area of your ***census enumeration village*** on this map.

    [RA NOTE] Please be as precise as possible in drawing the boundary. Use a **BLACK MARKER or PEN**.

23. **[SA]** What are some reasons why you defined your ***census enumeration village*** boundaries like this?

    [RA NOTE] Please provide this response in ENGLISH

    Response in ENGLISH

24. **[SA]** What is the closest landmark to your home? By landmark, we mean a location that most people in your ***census enumeration village*** would recognize.

    [RA NOTE] Please mark this location with a **GREEN CIRCLE** on your map.

    Notable landmark _________

25. **[BIN]** Do you belong to a smaller ***village*** within your ***census enumeration village***?

    ☐ No (SKIP to 27)
    ☐ Yes

26. **[MAP]** Please show me the boundary line of this smaller ***village*** on this map. Please be as precise as possible.

> **[RA NOTE]** Please be as precise as possible in drawing the boundary.

27. **[BIN]** Do you work ***within*** your ***census enumeration village*** boundaries?
(SKIP if 17 is NOT WORKING)

☐ No
☐ Yes

Activity Path Log

28. To begin, I will ask about all the places you visited yesterday, not only within your ***census enumeration village*** but anywhere surrounding this area. If you were not in your census enumeration village yesterday, please think of the most recent day you were in this area.

> **[RA NOTE]** *Identified times don't have to be exact; they can be approximate time points during the day.*

a. **[CAT]** Yesterday (or the day surveyed) was...

☐ Monday
☐ Tuesday
☐ Wednesday
☐ Thursday
☐ Friday
☐ Saturday
☐ Sunday

b. **[SA]** Start in the morning. First tell me ***where*** you woke up, ***when*** you left that place, and the ***first location*** you visited. On this map, show me where both of these locations are and how long you visited this ***first location***.

> **[RA NOTE]** Mark this ***wake-up*** location with a STAR ☆ using a **BLACK MARKER**. Mark this ***first location*** with a **number 1** using a **BLACK MARKER**.

c. **[SA]** On this map, please show the path you would take between the place you woke up and the first place you visited.

> **[RA NOTE]** While the participant points out the path, draw it on the map.

d. **[SA]** On this map, show me the next place you visited and for ***how long***. On the map, show the path that you'd take to this place.

> **[RA NOTE]** Mark this place with a **number 2** using a **BLACK MARKER**. While the participant points out the path, draw it on the map.

e. **[SA]** On this map, please go on throughout the entire day by showing all places you visited in sequential order, and the way you traveled from place to place.

> **[RA NOTE]** Mark each location with its corresponding number using a **BLACK MARKER**, continuing with as **many numbers as necessary**.

f. Daily Activity Table

[RA NOTE] Now, starting with the first place visited, indicate the departure location on this chart in ***prior location***. Please list these locations chronologically. Next, for each location visited, ask for the ***purpose*** of their travel, the ***mode of travel***, and ***with whom*** they traveled.

Variable Choices

Mode of travel
- 1. walk, 2. bike or bike taxi, 3. motorbike or motorbike taxi, 4. car, 5. other

With whom did you travel
- 1. alone, 2. not alone, with children, 3. not alone, without children

Time of day
- early morning: 4am-8am
- mid morning: 8am-10am
- late morning: 10am-12pm
- noon: 11:30am-12:30pm
- early afternoon: 12pm-2pm
- mid afternoon: 2pm-4pm
- late afternoon: 4pm-6pm
- evening: 6pm-12am
- twilight: 12am-4am

Loc #	Time of day	Location Visited	Prior Location	Time at location	Purpose	Mode of travel	Travel time	With whom did you travel
1.								
2.								
3.								
4.								
5.								

29. **[MAP]** Are there any locations you visited yesterday (or the day surveyed) that you would ***not*** normally visit?

☐ No (SKIP to 31)
☐ Yes

a. **[SA]** On the map, please show me these locations so I can mark them.

> **[RA NOTE]** Use the **RED MARKER** to circle these locations on the map with a ◯.

30. Are there any locations you ***did not*** visit yesterday (or the day surveyed) that you would normally visit?

☐ No (SKIP to 32)
☐ Yes

a. **[SA]** On the map, please show me these locations so I can mark them.

> **[RA NOTE]** Use the **BLUE MARKER** to mark these locations on the map with **numbers**.

b. **[SA]** On the map, show the path of how you would normally travel to these locations from a location that you would normally depart from.

> **[RA NOTE]** Please mark departure locations with the **BLUE MARKER** with "**[NUMBER]s**". For example, for the starting location 1, the departure location will be 1s.

c. Missed Frequently Visited Locations Activity Table

[RA NOTE] Fill in the following chart with the listed locations from part (a) with their corresponding departure point. For each trip, indicate the ***purpose of this travel***, the ***mode of travel***, and with ***whom you would normally travel***.

Variable Choices
Mode of travel
- 1. walk, 2. bike or bike taxi, 3. motorbike or motorbike taxi, 4. car, 5. other

With whom did you travel
- 1. alone, 2. alone, with children, 3. alone, without children

Location	Prior Location	Purpose of Travel	Mode of travel	With whom did you travel

31. **[SA]** Are there places that you visit to have fun or find enjoyment within your ***census enumeration village*** boundaries? (e.g. playing football, bawo, watching football, village dances)

☐ No (SKIP to END)
☐ Yes

a. **[SA]** On the map, please show me these locations so I can mark them.

> **[RA NOTE]** Please draw a number enclosed with a triangle △ using a **BLACK MARKER** around these locations on their map.

b. **[SA]** On the map, show the path of how you would normally travel to these locations from a location that you would normally depart from.

> **[RA NOTE]** Please mark departure locations with the **BLACK MARKER** with "[**NUMBER]s"** enclosed with a triangle. For example, for the starting location 1, the departure location will be 1s, both enclosed in a triangle.

c. Leisure Location Table

[RA NOTE] Fill in the following chart with the listed locations from part (a) with their corresponding departure point. For each trip, indicate the ***purpose of this travel***, the ***mode of travel***, ***with whom you would normally travel****, and the* ***day of the week*** *you would normally visit this location*.

Variable Choices
Mode of travel

- 1. walk, 2. bike or bike taxi, 3. motorbike or motorbike taxi, 4. car, 5. other

With whom did you travel

- 1. alone, 2. alone, with children, 3. alone, without children

Days of the week

- Monday to Sunday, *list all that apply*

Location	Prior Location	Purpose of Travel	Mode of travel	With whom did you travel	Days of the week

32. **[GEO]** Use SurveyCTO to capture the geocode of the participant's home.

[RA NOTE] *Please do this by first moving to their front door. If you are talking with the participant away from his/her home, please move back to his/her front door to capture their household location.*

If the tablet or SurveyCTO is not working, use your mobile phone to capture a geocode of the participant's home.

Appendix C

Example Trajectory Response Map (Scan)


pentecostal church
JA10245

Bibliography

[1] World Bank Group. Measuring Rural Access: Using New Technologies. Working Paper, World Bank, Washington, DC (2016). URL https://openknowledge.worldbank.org/handle/10986/25187. Accepted: 2016-10-18T16:42:01Z.

[2] World Bank Group. Measuring Rural Access. Tech. Rep. ACS26526, World Bank (2019).

[3] LDC Category | Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and Small Island Developing States. URL https://www.un.org/ohrlls/content/ldc-category.

[4] Sustainable Development Solutions Network: A Global Initiative for the United Nations. Sustainable Development Goal 11.2. URL https://indicators.report/targets/11-2/.

[5] Mwale, M., Luke, R. & Pisa, N. Factors that affect travel behaviour in developing cities: A methodological review. *Transportation Research Interdisciplinary Perspectives* **16**, 100683 (2022). URL https://www.sciencedirect.com/science/article/pii/S2590198222001439.

[6] World Bank. Rural population (% of total population) - World. URL https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS.

[7] Bank, W. Rural population (% of total population) - Low income. URL https://data.worldbank.org.

[8] Sprague, C. & Simon, S. E. Ending HIV in the USA: integrating social determinants of health. *The Lancet* **398**, 742–743 (2021). URL https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)01236-8/fulltext. Publisher: Elsevier.

[9] Nijman, J. A Study of Space in Mumbai's Slums. *Tijdschrift voor economische en sociale geografie* **101**, 4–17. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9663.2009.00576.x.

[10] Evans, T., Whitehead, M. & Diderichsen, F. *Challenging Inequities in Health: From Ethics to Action* (Oxford University Press, 2001). Google-Books-ID: XYYRDAAAQBAJ.

[11] Ezeh, A. *et al.* The history, geography, and sociology of slums and the health problems of people who live in slums. *The Lancet* **389**, 547–558 (2017). URL `http://www.sciencedirect.com/science/article/pii/S0140673616316506`.

[12] Marmot, M. Social determinants of health inequalities. *The Lancet* **365**, 1099–1104 (2005). URL `http://www.sciencedirect.com/science/article/pii/S0140673605711466`.

[13] Link, B. G. & Phelan, J. Social Conditions As Fundamental Causes of Disease. *Journal of Health and Social Behavior* **35**, 80 (1995). URL `http://www.jstor.org/stable/2626958?origin=crossref`.

[14] HIV and AIDS - Basic facts. URL `https://www.unaids.org/en/frequently-asked-questions-about-hiv-and-aids`.

[15] Singh, G. K. & Siahpush, M. Widening Rural–Urban Disparities in All-Cause Mortality and Mortality from Major Causes of Death in the USA, 1969–2009. *Journal of Urban Health* **91**, 272–292 (2014). URL `https://doi.org/10.1007/s11524-013-9847-2`.

[16] Hartley, D. Rural Health Disparities, Population Health, and Rural Culture. *American Journal of Public Health* **94**, 1675–1678 (2004). URL `https://ajph.aphapublications.org/doi/full/10.2105/AJPH.94.10.1675`. Publisher: American Public Health Association.

[17] Strasser, R., Kam, S. M. & Regalado, S. M. Rural Health Care Access and Policy in Developing Countries. *Annual Review of Public Health* **37**, 395–412 (2016). URL `https://doi.org/10.1146/annurev-publhealth-032315-021507`. _eprint: https://doi.org/10.1146/annurev-publhealth-032315-021507.

[18] Access to electricity – SDG7: Data and Projections – Analysis. URL `https://www.iea.org/reports/sdg7-data-and-projections/access-to-electricity`.

[19] Water and Sanitation - United Nations Sustainable Development. URL `https://www.un.org/sustainabledevelopment/water-and-sanitation/`.

[20] Khan, M. H. Rural Poverty in Developing Countries: Implications for Public Policy. URL `https://www.imf.org/en/Publications/Economic-Issues/Issues/2016/12/30/Rural-Poverty-in-Developing-Countries-Implications-for-Public-Policy-14783`.

[21] Yaya, S., Bishwajit, G. & Shah, V. Wealth, education and urban–rural inequality and maternal healthcare service usage in Malawi. *BMJ Global Health* **1**, e000085 (2016). URL https://gh.bmj.com/lookup/doi/10.1136/bmjgh-2016-000085.

[22] United Nations. Tackling inequalities in public service coverage to “build forward better” for the rural poor. Tech. Rep., United Nations, New York, NY: United Nations (2021).

[23] How do we define cities, towns, and rural areas? (2020). URL https://blogs.worldbank.org/sustainablecities/how-do-we-define-cities-towns-and-rural-areas.

[24] Gee, G. C. & Payne-Sturges, D. C. Environmental Health Disparities: A Framework Integrating Psychosocial and Environmental Concepts. *Environmental Health Perspectives* **112**, 1645–1653 (2004). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1253653/.

[25] Kawachi, I., Subramanian, S. V. & Almeida-Filho, N. A glossary for health inequalities. *Journal of Epidemiology & Community Health* **56**, 647–652 (2002). URL https://jech.bmj.com/content/56/9/647.

[26] World Bank. Malawi 2022 Population. URL https://data.worldbank.org.

[27] UNAIDS. Malawi HIV Country Fact Sheet 2022. URL https://www.unaids.org/en/regionscountries/countries/malawi.

[28] Frescura, L. *et al.* Achieving the 95 95 95 targets for all: A pathway to ending AIDS. *PLoS ONE* **17**, e0272405 (2022). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9352102/.

[29] Libamba, E. *et al.* Supervision, monitoring and evaluation of nationwide scale-up of antiretroviral therapy in Malawi. *Bulletin of the World Health Organization* **84**, 320–326 (2006). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2627329/.

[30] Harries, A. D., Makombe, S. D., Libamba, E. & Schouten, E. J. Why did the scale-up of HIV treatment work? A case example from Malawi. *Journal of Acquired Immune Deficiency Syndromes (1999)* **57 Suppl 2**, S64–67 (2011).

[31] Jahn, A. *et al.* Scaling-up antiretroviral therapy in Malawi. *Bulletin of the World Health Organization* **94**, 772–776 (2016). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5043204/.

[32] Masefield, S. C., Msosa, A. & Grugel, J. Challenges to effective governance in a low income healthcare system: a qualitative study of stakeholder perceptions in Malawi. *BMC health services research* **20**, 1142 (2020).

[33] Wolock, T. M. *et al.* Subnational HIV incidence trends in Malawi: large, heterogeneous declines across space. *medRxiv* 2023.02.02.23285334 (2023). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9915821/.

[34] Addressing a blind spot in the response to HIV — Reaching out to men and boys. URL https://www.unaids.org/en/resources/documents/2017/blind_spot.

[35] Hensen, B., Taoka, S., Lewis, J. J., Weiss, H. A. & Hargreaves, J. Systematic review of strategies to increase men's HIV-testing in sub-Saharan Africa. *AIDS (London, England)* **28**, 2133–2145 (2014).

[36] Horton, K. C., MacPherson, P., Houben, R. M. G. J., White, R. G. & Corbett, E. L. Sex Differences in Tuberculosis Burden and Notifications in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. *PLoS medicine* **13**, e1002119 (2016).

[37] World Bank. Rural population (% of total population) - Malawi. URL https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS?locations=MW.

[38] Oxford Poverty and Human Development Initiative. URL https://ophi.org.uk/multidimensional-poverty-index/mpi-country-briefings/.

[39] Beegle, K., Christiaensen, L., Dabalen, A. & Gaddis, I. *Poverty in a Rising Africa* (World Bank Publications, 2016). Google-Books-ID: zADYCwAAQBAJ.

[40] Malawi Roads Authority. URL https://www.ra.org.mw/quick-stats/.

[41] Diez Roux, A. V. Investigating neighborhood and area effects on health. *American Journal of Public Health* **91**, 1783–1789 (2001).

[42] Macintyre, S., Ellaway, A. & Cummins, S. Place effects on health: how can we conceptualise, operationalise and measure them? *Social Science & Medicine* **55**, 125–139 (2002). URL http://www.sciencedirect.com/science/article/pii/S0277953601002143.

[43] Msisha, W. M., Kapiga, S. H., Earls, F. J. & Subramanian, S. V. Place matters: multilevel investigation of HIV distribution in Tanzania. *AIDS* **22**, 741 (2008). URL https://journals.lww.com/aidsonline/Fulltext/2008/03300/Place_matters__multilevel_investigation_of_HIV.00009.aspx?casa_token=d3tGNrAePPwAAAAA:bb0g-GmLfPprrclPXi6RHgggZjqz20CAKVMS5EIpRjxAr01-RZUVn8fK975Yhsjm3DkSSffXGjw8KXzI

[44] Figueroa, C., Johnson, C., Verster, A. & Baggaley, R. Attitudes and Acceptability on HIV Self-testing Among Key Populations: A Literature Review. *AIDS and Behavior* **19**, 1949–1965 (2015). URL https://doi.org/10.1007/s10461-015-1097-8.

[45] Kayeyi, N., Fylkesnes, K., Michelo, C., Makasa, M. & Sandøy, I. Decline in HIV Prevalence among Young Women in Zambia: National-Level Estimates of Trends Mask Geographical and Socio-Demographic Differences. *PLOS ONE* **7**, e33652 (2012). URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0033652. Publisher: Public Library of Science.

[46] Boerma, J. T., Urassa, M., Senkoro, K., Klokke, A. & Ngweshemi, J. Z. L. Spread of HIV infection in a rural area of Tanzania. *AIDS* **13**, 1233 (1999). URL https://journals.lww.com/aidsonline/fulltext/1999/07090/spread_of_hiv_infection_in_a_rural_area_of.13.aspx?casa_token=5ndNyJcwYmAAAAAA:eYcRFAu8nBTnVVKxojQpxGG3WOHUWP42fsCFoCffKIHxit7yY-euYrxCKhZqTIOkTAbt2kIa94fKYJde

[47] Drame, F. M. & Foley, E. E. HIV/AIDS in mid-sized cities in Senegal: From individual to place-based vulnerability. *Social Science & Medicine* **133**, 296–303 (2015). URL https://www.sciencedirect.com/science/article/pii/S0277953614007709.

[48] Tempalski, B. & McQuie, H. Drugscapes and the role of place and space in injection drug use-related HIV risk environments. *International Journal of Drug Policy* **20**, 4–13 (2009). URL https://www.sciencedirect.com/science/article/pii/S0955395908000339.

[49] Zenk, S. N. *et al.* Activity space environment and dietary and physical activity behaviors: a pilot study. *Health & Place* **17**, 1150–1161 (2011).

[50] Matthews, S. A. & Yang, T.-C. Spatial Polygamy and Contextual Exposures (SPACEs): Promoting Activity Space Approaches in Research on Place And Health. *American Behavioral Scientist* **57**, 1057–1081 (2013). URL http://journals.sagepub.com/doi/10.1177/0002764213487345.

[51] Perchoux, C., Chaix, B., Cummins, S. & Kestens, Y. Conceptualization and measurement of environmental exposure in epidemiology: Accounting for activity space related to daily mobility. *Health & Place* **21**, 86–93 (2013). URL https://www.sciencedirect.com/science/article/pii/S1353829213000117.

[52] Golledge, R. G. & Stimson, R. J. *Spatial Behavior: A Geographic Perspective* (The Guilford Press, New York, 1996).

[53] Conners, E. E. *et al.* Quantitative, Qualitative and Geospatial Methods to Characterize HIV Risk Environments. *PLOS ONE* **11**, e0155693 (2016). URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155693. Publisher: Public Library of Science.

[54] Theall, K. P. *et al.* Considering high alcohol and violence neighborhood context using daily diaries and GPS: A pilot study among people living with HIV. *Drug and Alcohol Dependence* **187**, 236–241 (2018). URL https://www.sciencedirect.com/science/article/pii/S0376871618301893.

[55] Arnold, T. *et al.* Locating the Risk: Using Participatory Mapping to Contextualize Perceived HIV Risk across Geography and Social Networks among Men Who Have Sex with Men in the Deep South. *The Journal of Sex Research* **59**, 931–938 (2022). URL https://doi.org/10.1080/00224499.2021.1906397. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00224499.2021.1906397.

[56] Davoudi, S. & Stead, D. Urban-Rural Relationships: An Introduction and Brief History. *Built Environment (1978-)* **28**, 268–277 (2002).

[57] Cummins, S., Curtis, S., Diez-Roux, A. V. & Macintyre, S. Understanding and representing 'place' in health research: A relational approach. *Social Science & Medicine* **65**, 1825–1838 (2007). URL https://www.sciencedirect.com/science/article/pii/S0277953607003061.

[58] Litosseliti, L. *Using Focus Groups in Research* (A&C Black, 2003). Google-Books-ID: GwLbQSlRtQAC.

[59] Leung, F.-H. & Savithiri, R. Spotlight on focus groups. *Canadian Family Physician* **55**, 218–219 (2009). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2642503/.

[60] Kitzinger, J. Qualitative Research: Introducing focus groups. *BMJ* **311**, 299–302 (1995). URL https://www.bmj.com/content/311/7000/299. Publisher: British Medical Journal Publishing Group Section: Education and debate.

[61] Creswell, J. W. & Clark, V. L. P. *Designing and Conducting Mixed Methods Research* (SAGE Publications, 2017). Google-Books-ID: BXEzDwAAQBAJ.

[62] Felker-Kantor, E. *et al.* Participatory geographic mapping and activity space diaries: innovative data collection methods for understanding environmental risk exposures among female sex workers in a low-to middle-income country. *International Journal of Health Geographics* **20**, 25 (2021). URL https://doi.org/10.1186/s12942-021-00279-9.

[63] Cochrane, L. & Corbett, J. Participatory Mapping. In Servaes, J. (ed.) *Handbook of Communication for Development and Social Change*, 705–713 (Springer, Singapore, 2020). URL https://doi.org/10.1007/978-981-15-2014-3_6.

[64] Fang, M. L. *et al.* Place-making with older persons: Establishing sense-of-place through participatory community mapping workshops. *Social Science & Medicine* **168**, 223–229 (2016). URL https://www.sciencedirect.com/science/article/pii/S0277953616303616.

[65] Good practices in participatory mapping: A review prepared for the International Fund for Agricultural Development (IFAD) (2009).

[66] Axhausen, K. & Zürich, E. Concepts of Travel Behavior Research (2007).

[67] Pedisic, Z., Dumuid, D. & Olds, T. S. Integrating Sleep, Sedentary Behaviour, And Physical Activity Research In The Emerging Field Of Time-Use Epidemiology: Definitions, Concepts, Statistical Methods, Theoretical Framework, And Future Directions (2017).

[68] Takahashi, L. M., Wiebe, D. & Rodriguez, R. Navigating the time–space context of HIV and AIDS:: daily routines and access to care. *Social Science & Medicine* **53**, 845–863 (2001). URL https://www.sciencedirect.com/science/article/pii/S0277953600003634.

[69] Broeg, W., Fallast, K., Katteler, H., Sammer, G. & Schwertner, B. Selected results of a standardized survey instrument for large-scale travel surveys in several european countries. *Publication of: VNU SCIENCE PRESS* (1985). URL https://trid.trb.org/view/272438.

[70] Ampt, E. S., Richardson, A. J. & Brög, W. *New Survey Methods in Transport: Proceedings of 2nd International Conference, Hungerford Hill, Australia, 12-16 September 1983* (VSP, 1985). Google-Books-ID: nvYSL9JPB00C.

[71] Priya Uteng, T. & Turner, J. Addressing the Linkages between Gender and Transport in Low- and Middle-Income Countries. *Sustainability* **11**, 4555 (2019). URL https://www.mdpi.com/2071-1050/11/17/4555. Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.

[72] Hidayati, I., Tan, W. & Yamu, C. Conceptualizing Mobility Inequality: Mobility and Accessibility for the Marginalized. *Journal of Planning Literature* **36**, 492–507 (2021). URL https://doi.org/10.1177/08854122211012898. Publisher: SAGE Publications Inc.

[73] Dumuid, D. *et al.* Compositional Data Analysis in Time-Use Epidemiology: What, Why, How. *International Journal of Environmental Research and Public Health* **17**, 2220 (2020). URL https://www.mdpi.com/1660-4601/17/7/2220. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

[74] Paffenbarger, R. S., Jr, Blair, S. N. & Lee, I.-M. A history of physical activity, cardiovascular health and longevity: the scientific contributions of Jeremy N Morris, DSc, DPH, FRCP. *International Journal of Epidemiology* **30**, 1184–1192 (2001). URL https://doi.org/10.1093/ije/30.5.1184.

[75] Jette, M., Sidney, K. & Blümchen, G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clinical Cardiology* **13**, 555–565 (1990).

[76] Ainsworth, B. *et al.* 2011 compendium of physical activities: A second update of codes and MET values. *Medicine and Science in Sports and Exercise* **43**, 1575–1581 (2011). URL http://www.scopus.com/inward/record.url?scp=79961091187&partnerID=8YFLogxK.

[77] What is Sedentary Behaviour? URL https://www.sedentarybehaviour.org/what-is-sedentary-behaviour/.

[78] Owen, N., Healy, G. N., Matthews, C. E. & Dunstan, D. W. Too Much Sitting: The Population-Health Science of Sedentary Behavior. *Exercise and Sport Sciences Reviews* **38**, 105–113 (2010). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3404815/.

[79] Pate, R. R., O'Neill, J. R. & Lobelo, F. The Evolving Definition of "Sedentary". *Exercise and Sport Sciences Reviews* **36**, 173 (2008). URL https://journals.lww.com/acsm-essr/Fulltext/2008/10000/The_Evolving_Definition_of__Sedentary_.00002.aspx?casa_token=Cw9kCcBc3zcAAAAA:Z6XrSA2xq_TRTNoOuYR5xlytqzdoq7Mjuakw_5HZDC_MQ5LlFQoR-tvyTMZL9KCw6zP12dL3FHkg5tGDPSBsez4.

[80] Tremblay, M. S., Colley, R. C., Saunders, T. J., Healy, G. N. & Owen, N. Physiological and health implications of a sedentary lifestyle. *Applied Physiology, Nutrition, and Metabolism* **35**, 725–740 (2010). URL https://cdnsciencepub.com/doi/full/10.1139/H10-079. Publisher: NRC Research Press.

[81] Tremblay, M. S. *et al.* New Canadian Physical Activity Guidelines. *Applied Physiology, Nutrition, and Metabolism* **36**, 36–46 (2011). URL https://cdnsciencepub.com/doi/full/10.1139/H11-009. Publisher: NRC Research Press.

[82] Kwan, M.-P. The Uncertain Geographic Context Problem. *Annals of the Association of American Geographers* **102**, 958–968 (2012). URL https://doi.org/10.1080/00045608.2012.687349.

[83] Laube, P. *Computational Movement Analysis*. SpringerBriefs in Computer Science (Springer International Publishing, Cham, 2014). URL `https://link.springer.com/10.1007/978-3-319-10268-9`.

[84] Mazimpaka, J. D. & Timpf, S. Trajectory data mining: A review of methods and applications. *Journal of Spatial Information Science* 61–99 (2016). URL `https://josis.org/index.php/josis/article/view/73`.

[85] Kong, X. *et al.* Big Trajectory Data: A Survey of Applications and Services. *IEEE Access* **6**, 58295–58306 (2018).

[86] Wang, D., Miwa, T. & Morikawa, T. Big Trajectory Data Mining: A Survey of Methods, Applications, and Services. *Sensors (Basel, Switzerland)* **20**, 4571 (2020). URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7472055/`.

[87] Lin, D. An Information-Theoretic Definition of Similarity .

[88] Faloutsos, C., Jagadish, H., Mendelzon, A. & Milo, T. A signature technique for similarity-based queries. In *Proceedings. Compression and Complexity of SEQUENCES 1997 (Cat. No.97TB100171)*, 2–20 (1997).

[89] Ding, H., Trajcevski, G., Scheuermann, P., Wang, X. & Keogh, E. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment* **1**, 1542–1552 (2008). URL `https://doi.org/10.14778/1454159.1454226`.

[90] The Computational Geometry of Comparing Shapes | SpringerLink. URL `https://link.springer.com/chapter/10.1007/978-3-642-03456-5_16`.

[91] Demsar, U. *et al.* Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology* **3**, 5 (2015).

[92] Cleasby, I. R. *et al.* Using time-series similarity measures to compare animal movement trajectories in ecology. *Behavioral Ecology and Sociobiology* **73**, 151 (2019). URL `https://doi.org/10.1007/s00265-019-2761-1`.

[93] Senin, P. Dynamic Time Warping Algorithm Review .

[94] Wang, H., Su, H., Zheng, K., Sadiq, S. & Zhou, X. An effectiveness study on trajectory similarity measures. *Database Technologies* **137** (2013).

[95] Vlachos, M., Kollios, G. & Gunopulos, D. Discovering similar multidimensional trajectories. In *Proceedings 18th International Conference on Data Engineering*, 673–684 (2002). ISSN: 1063-6382.

[96] Chen, L., Ozsu, M. T. & Oria, V. Robust and fast similarity search for moving object trajectories. *Proceedings of the ACM SIGMOD International Conference on Management of Data* 491–502 (2005).

[97] Fréchet, M. M. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884-1940)* **22**, 1–72 (1906). URL https://doi.org/10.1007/BF03018603.

[98] Toohey, K. & Duckham, M. Trajectory similarity measures. *SIGSPATIAL Special* **7**, 43–50 (2015). URL https://doi.org/10.1145/2782759.2782767.

[99] SurveyCTO. URL https://www.surveycto.com/.

[100] FANTA-2 Project. Alternative Sampling Designs for Emergency Settings: A Guide for Survey Planning, Data Collection and Analysis (2009). URL https://www.fsnnetwork.org/sites/default/files/alternative_sampling_designs_for_emergency_settings.pdf.

[101] Aitchison, J. The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological)* **44**, 139–177 (1982). URL https://www.jstor.org/stable/2345821. Publisher: [Royal Statistical Society, Wiley].

[102] Gupta, N. *et al.* A comparison of standard and compositional data analysis in studies addressing group differences in sedentary behavior and physical activity. *International Journal of Behavioral Nutrition and Physical Activity* **15**, 53 (2018). URL https://doi.org/10.1186/s12966-018-0685-1.

[103] Chastin, S. F. M., Palarea-Albaladejo, J., Dontje, M. L. & Skelton, D. A. Combined Effects of Time Spent in Physical Activity, Sedentary Behaviors and Sleep on Obesity and Cardio-Metabolic Health Markers: A Novel Compositional Data Analysis Approach. *PLOS ONE* **10**, e0139984 (2015). URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0139984. Publisher: Public Library of Science.

[104] Pedisic, Z. Measurement issues and poor adjustments for physical activity and sleep undermine sedentary behaviour research—The focus should shift to the balance between sleep, sedentary behaviour, standing and activity. *Kinesiology* **46**, 135–146 (2014).

[105] Tein, J.-Y., Coxe, S. & Cham, H. Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. *Structural Equation Modeling: A Multidisciplinary Journal* **20**, 640–657 (2013). URL https://doi.org/10.1080/10705511.2013.824781. Publisher: Routledge _eprint: https://doi.org/10.1080/10705511.2013.824781.

[106] Berlin, K. S., Williams, N. A. & Parra, G. R. An Introduction to Latent Variable Mixture Modeling (Part 1): Overview and Cross-Sectional Latent Class and Latent Profile Analyses. *Journal of Pediatric Psychology* **39**, 174–187 (2014). URL `https://doi.org/10.1093/jpepsy/jst084`.

[107] Gupta, N. *et al.* Movement behavior profiles and obesity: a latent profile analysis of 24-h time-use composition among Danish workers. *International Journal of Obesity* **44**, 409–417 (2020). URL `https://www.nature.com/articles/s41366-019-0419-8`. Number: 2 Publisher: Nature Publishing Group.

[108] Sardá-Espinosa, A. Time-Series Clustering in R Using the dtwclust Package. *The R Journal* **11**, 22 (2019). URL `https://journal.r-project.org/archive/2019/RJ-2019-023/index.html`.

[109] Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of Predictability in Human Mobility. *Science* **327**, 1018–1021 (2010). URL `https://www.science.org/doi/10.1126/science.1177170`.

[110] González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008). URL `https://www.nature.com/articles/nature06958`.

[111] Kang, C., Ma, X., Tong, D. & Liu, Y. Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and its Applications* **391**, 1702–1717 (2012). URL `https://ideas.repec.org//a/eee/phsmap/v391y2012i4p1702-1717.html`. Publisher: Elsevier.

[112] Yuan, Y. & Raubal, M. Analyzing the distribution of human activity space from mobile phone usage: an individual and urban-oriented study. *International Journal of Geographical Information Science* **30**, 1594–1621 (2016). URL `https://www.tandfonline.com/doi/full/10.1080/13658816.2016.1143555`.

[113] Xu, Y., Belyi, A., Bojic, I. & Ratti, C. Human mobility and socioeconomic status: Analysis of Singapore and Boston. *Computers, Environment and Urban Systems* **72**, 51–67 (2018). URL `https://linkinghub.elsevier.com/retrieve/pii/S0198971517304179`.

[114] Pappalardo, L., Rinzivillo, S., Qu, Z., Pedreschi, D. & Giannotti, F. Understanding the patterns of car travel. *The European Physical Journal Special Topics* **215**, 61–73 (2013). URL `http://link.springer.com/10.1140/epjst/e2013-01715-5`.

[115] Sarda-Espinosa, A. dtwclust: Time Series Clustering Along with Optimizations for the Dynamic Time Warping Distance (2022). URL https://CRAN.R-project.org/package=dtwclust.

[116] Kaufman, L. & Rousseeuw, P. *Finding Groups in Data: An Introduction To Cluster Analysis* (1990). Journal Abbreviation: Wiley, New York. ISBN 0-471-87876-6. Publication Title: Wiley, New York. ISBN 0-471-87876-6.

[117] Pappalardo, L., Simini, F., Barlacchi, G. & Pellungrini, R. scikit-mobility (2019).

[118] Ha, J. H. *et al.* Gendered relationship between HIV stigma and HIV testing among men and women in Mozambique: a cross-sectional study to inform a stigma reduction and male-targeted HIV testing intervention. *BMJ Open* **9**, e029748 (2019). URL https://bmjopen.bmj.com/content/9/10/e029748. Publisher: British Medical Journal Publishing Group Section: HIV/AIDS.

[119] Fonner, V. A., Mbwambo, J. K., Kennedy, C. E. & Sweat, M. D. The gendered experience of HIV testing: factors associated with prior testing differ among men and women in rural Tanzania. *International journal of STD & AIDS* **30**, 843–852 (2019). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6737529/.

[120] Hlongwa, M., Mashamba-Thompson, T., Makhunga, S., Muraraneza, C. & Hlongwana, K. Men's perspectives on HIV self-testing in sub-Saharan Africa: a systematic review and meta-synthesis. *BMC Public Health* **20**, 66 (2020). URL https://doi.org/10.1186/s12889-020-8184-0.