

https://dspace.org/

An integrated knowledge ecosystem for infant mortality in Kilifi health and demographic surveillance system (KHDSS)

Amadi, David Kivuli

2019
https://hdl.handle.net/10539/29716

Downloaded from a DSpace repository. https://dspace.org/

An Integrated Knowledge Ecosystem for Infant Mortality in Kilifi Health and Demographic Surveillance System (KHDSS)



DAVID KIVULI AMADI

STUDENT NUMBER: 969197

A Research Report submitted to the Faculty of Health Science in partial fulfillment of the requirements for the degree of *Master of Science (MSc)* in Epidemiology - Public Health Informatics

16th October, 2019

David Kivuli Amadi . *2018*
An Integrated Knowledge Ecosystem for Infant Mortality in Kilifi Health and Demographic Surveillance System (KHDSS).

Copyright © University of the Witwatersrand, Johannesburg, South Africa.

All rights reserved. No part of this research report may be stored in a retrieval system, transmitted, or reproduced, in any form or by any means, including but not limited to photocopy, photograph, magnetic or other record, without prior agreement and written permission of the copyright holder.

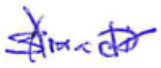
SUPERVISORS:
Dr. Gideon Nimako, University of The Witwatersrand
Mr. Michael Mapundu, University of The Witwatersrand

SUPPORTED BY:
University of the Witwatersrand
KEMRI | wellcome trust-Kilifi

DECLARATION

I hereby declare that this research report is project work carried out by me under the guidance of Dr. Gideon Nimako and Mr. Michael Mapundu. I have taken care in all respects to honour the intellectual property rights of others and have acknowledged the contribution of others. I further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Parktown, Johannesburg, 16th *October,* 2019



David Kivuli Amadi

DEDICATION

I would like to dedicate my thesis to all the Amadi family members. Special dedication to my late dad, Mr. P.A.A who inspired and prepared me to face challenges with faith and humility not forgetting his favorite quote 'Live within your means'. Thanks to my wife Beth and Son Dylan and ever praying mom Nabash Nyasaye akugasise na akumede mihiga minyingi.

ABSTRACT

Infant mortality is a key population indicator and remains a problem that requires global attention. Kenya, like many countries in Sub Saharan Africa, did not achieve the Millennium Development Goal number 4 (Reducing childhood mortality by 2015). Factors that contribute to infant mortality require scientific evidence thus focus has been put on visualization applications for the generation of new hypotheses and to inform decision. The need for rapid access to information in an explosive data generation in the public health sector needs a more effective data automation procedure.

The aim of this project is to develop a digital ecosystem platform that brings together data from selected observational studies, randomized controlled clinical trials, and national surveys into large composite datasets that give analytical power to answer important questions on child health and surface the insights that create impact.

Using 2015 data from Kilifi HDSS for children aged less than one year. Shiny data-driven web application framework for R statistical computing explores and analyses data on a dashboard to run algorithms. The platform provides utilities for data analytics and modeling techniques to answer questions on child health such as the relative effects of pre and postnatal impacts on physical growth.

Exploratory data analysis, descriptive statistics as well as multivariable analysis are presented by the use of graphs, tables, maps and rate(s) ratios. The visualization platform allows researchers and policymakers to generate actionable recommendations, predictions, and new hypotheses that will inform decision.

ACKNOWLEDGMENTS

I wish to acknowledge the following people and organisations:

- I would like to express my appreciation to Dr. Gideon Nimako and Mr. Michael Mapundu for their guidance and supervision.
- Sincere thanks to my family for their support and patience throughout the study period.
- I would like to express my gratitude towards Kemri Wellcome Trust Epidemiology department for allowing me use their data.
- All the lecturers at the Division of Epidemiology and Biostatistics at Wits School of Public Health.
- Finally, thanks to INDEPTH Network for awarding me the MSc scholarship.

CONTENTS

DECLARATION iii
DEDICATION iv
ABSTRACT v
ACKNOWLEDGMENTS vi
ACRONYMS ix
List of Figures x
List of Tables xi
1 INTRODUCTION 1
1.1 Problem Statement . . . 3
1.2 Motivation . . . 3
1.3 Research Aim and Objectives . . . 4
1.4 Outline of Research Report . . . 4
2 LITERATURE REVIEW 5
2.1 Overview of Infant Mortality . . . 5
2.2 Maternal Health Factors . . . 6
2.3 Socioeconomic Determinants . . . 7
2.4 Immunization . . . 8
2.5 Child Health Factors . . . 9
2.6 Visualization . . . 10
2.7 Visual Analytics Applications . . . 11
2.8 Reproducible Research . . . 12
2.9 Summary of the Literature . . . 13
3 DATA MODEL FOR THE INTEGRATED KNOWLEDGE ECOSYSTEM 14
3.1 Introduction . . . 14
3.2 Data Sources . . . 14
3.3 Kilifi HDSS Data Model . . . 15
3.4 Description of Core Data . . . 16
3.5 Variables and Variable Measurements . . . 17
3.5.1 Dependent Variables . . . 17
3.5.2 Independent Variables . . . 17
3.6 Statistical Analysis . . . 17
3.7 Overview of the technology used . . . 19
3.8 R (Programming Language) . . . 19
3.9 R Shiny . . . 19
3.10 R MarkDown . . . 20
3.11 Dependencies . . . 20
3.12 Architecture . . . 20
3.13 Model . . . 22
3.14 summary . . . 23
4 UTILITY TOOLS AND MODULES FOR THE INTEGRATED KNOWLEDGE ECOSYSTEM 24
4.1 Introduction . . . 24

4.2 Dataset Tab . . . 24
4.3 Infographics Tab . . . 26
4.4 Graphics Tab . . . 27
4.5 Exploratory Analysis . . . 28
4.6 Regression Tab . . . 29
4.7 Trends Tab . . . 29
4.8 Spatial Distribution Tab . . . 31
4.9 Results . . . 32
4.10 Summary . . . 34
5 USING THE INTEGRATED KNOWLEDGE ECOSYSTEM IN OTHER HDSS 35
5.1 A Situational Analysis of the Current Data-Ecosystem Context of HDSSs and their Relation to the Platform . . . 35
5.2 Deployment consideration and Data Linkage . . . 38
5.3 Infrastructure consideration . . . 40
5.4 Summary . . . 42
6 CONCLUSION AND FUTURE DIRECTIONS 43
6.1 Conclusion . . . 43
6.2 Future Direction . . . 44
A PLAGIARISM DECLARATION 46
B ETHICS CLEARANCE CERTIFICATE 47
C DEPLOYMENT OF THE PLATFORM 48

BIBLIOGRAPHY 49

ACRONYMS

ANC Antenatal Care

BMGF Bill and Melinda Gates Foundation

CDC Centers for Disease Control

CSS Cascading Style Sheets

DHS Demographic and Health Survey

DGC Data Governance Committee

GIS Geographical Information System

DHS Demographic and Health Survey

INDEPTH International Network for the Demographic Evaluation of Populations and their Health

IDMP INDEPTH Data Management Programme

ETL Extract Transform and Load

GIS Geographic Information Systems

GLM Generalized linear model

GIVS Global Immunization Vision and Strategy

GSK GlaxoSmithKline

HDSS Health and Demographic Surveillance System

KDHS Kenya Demographic and Health Survey

KHDSS Kilifi Health and Demographic Surveillance System

KIDMS Kilifi Integrated Data Management System

MDG Millennium Development Goals

MHC Maternal Health Care

MOH Ministry of Health

RDBMS Relational Database Management System

RDM Reference Demographic Surveillance Data Model

OS Operating System

SDG Sustainable Development Goals

UNICEF United Nations International Children's Emergency Fund

USA United States of America

VMS Vaccine Monitoring System

WHO World Health Organization

LIST OF FIGURES

Figure 1 Kilifi Health and Demographic Surveillance System (HDSS) Data Model . . . 15
Figure 2 R Shiny Architecture . . . 21
Figure 3 Screenshot showing data upload sidebar with dataset variables, data structure, and summary statistics panels . . . 25
Figure 4 Screenshot showing KHDSS infographics . . . 26
Figure 5 Graphics tab display: A box plot of mother's age stratified by birth location . . . 27
Figure 6 Downloaded graph from the graphics tab display: A bar graph of immunization status stratified by birth location . . . 28
Figure 7 Screenshot showing tabulation tab . . . 29
Figure 8 Regression tab: Logistic regression display . . . 30
Figure 9 Trends Tab: KHDSS infants Death rates(2015) . . . 30
Figure 10 Spatial distribution of infant deaths in KHDSS . . . 31
Figure 11 Reference Demographic Surveillance Data Model . . . 36
Figure 12 Core data: Common event attributes for the INDEPTH data specification . . . 39
Figure 13 Health and Demographic Surveillance System linkage to Cohort studies . . . 40
Figure 14 Plagiarism Declaration . . . 46
Figure 15 Ethics Clearance Certificate . . . 47
Figure 16 Deployment of the Platform . . . 48

LIST OF TABLES

Table 1 Definition of variables 18
Table 2 Descriptive Characteristics 32
Table 3 Logistic regression analysis results 33
Table 4 Sample implementation at 6 HDSS sites 37
Table 5 Research studies conducted in International Network for the Demographic Evaluation of Populations and their Health (INDEPTH) member HDSSs in Africa 40

1

INTRODUCTION

The first year of a child's life is of special importance in health and plays a vital role in the development of national policies on health and remains a problem that requires a global attention [1]. Infant mortality remains high in sub-Saharan Africa despite the significant achievements made in the Millennium Development Goals (MDG) era. The MDG target 4 of two-thirds reduction of under 5 mortality was not achieved, this is evidenced by the decision to include the original MDG themes within the new health goal for the Sustainable Development Goals (SDG) by advising countries to prioritize child survival [2]. The SDG 2030 agenda has major implications for health monitoring by undertaking health data collection with great attention to disaggregated data, analysis and communicating in an integrated manner [3]. The determinants of infant mortality vary in Africa from socio-cultural to social-economic in comparison to the rest of the world [4]. Efforts to combat child mortality globally focuses on addressing inequities in access to care and public health interventions, however not much has been published [1].

Demographic and Health Survey (DHS) programs are an important source of data and provides information to help monitor and evaluate population and health status in developing countries. DHS survey data are widely acknowledged by researchers as data sources and provide key data for planning, monitoring and evaluating programs [5]. The Kilifi Health and Demographic Surveillance System (KHDSS), located in the Indian Ocean coast of Kenya, covers an area of $891 km^2$ comprised of 15 administrative locations, containing $280,000$ residents. Fieldworkers visit every participating household approximately three times a year to capture and record the incidence of pregnancies, births, deaths, and migration events. The site also supports multiple nested studies, acting as a framework for

epidemiological studies [6]. The Kilifi Integrated Data Management System (KIDMS) provides the data infrastructure for the demographic surveillance as well as research studies. Integration of the demographic surveillance system and research studies provides a huge potential for conducting health research and evaluating health interventions and outcomes.

The KHDSS supports activities aimed at evaluating new community-based interventions particularly preventing infectious diseases in children by use of vaccines. KHDSS data have been used to evaluate access to health care, testing association between genetic risk factors and infectious diseases, more importantly, it has been used to support the malaria vaccine clinical trial [6]. The GlaxoSmithKline (GSK) malaria vaccine, MosquirixTM RTS,S study conducted in Kilifi received positive opinion from European regulators and World Health Organization (WHO) will assess how the first malaria vaccine will be used to prevent malaria [7]. The department works closely with the Ministry of Health (MOH) to assist in planning and health care allocation as well. Further dissemination is achieved through publication of articles and sharing of data associated with the submission [8]. Longitudinal birth cohort studies where children and their families are followed up from birth to their current age allows tracking of growth, health well-being and the educational progress of their life.

The value of surveillance systems at KHDSS relies on the effective and efficient delivery of useful information. With the growth of surveillance systems and research studies at KHDSS coupled with a variety of health outcomes, emphasis needs to be placed for automated real-time analysis and dissemination of results via visual displays. The research findings generated by the automated tools can enable researchers to raise smart objectives and also allow health care providers to incorporate findings in their strategies and policies to improve health outcomes in a population.

1.1 PROBLEM STATEMENT

Researchers do not fully utilize the full potential of this flood of data from KHDSS. However, by simplifying the analytical approach they can exploit the potential in the data banks that will impact public health practice. This work provides a bridge to gap by developing an automated integrated analytic visualization tool as a baseline for a knowledge ecosystem to assist in the analysis of infant health data thereby expediting time spent on the analytical process.

1.2 MOTIVATION

The data collected by KIDMS applications is far more than its utilization for research purposes. Current child mortality analysis includes trends in mortality by sex, migration, and cause of death [9, 10]. Optimizing existing research studies data nested within KIDMS with core KHDSS variables that measure births, deaths, and migrations will provide valuable insights in explaining mortality. Furthermore, studies have shown that physicians who form a large proportion of scientists lack data analytical expertise due to insufficient training in statistics [11, 12]. Contrastingly, the need for statistical expertise by health care professionals has grown exponentially as healthcare becomes more of a data-driven endeavor [13]. With the current data culture, there is a need to invest in visual analytics tools and frameworks without the need for high technical or statistics background for end users. Such tools can aid researchers to design smart hypothesis on child and maternal health to address risk factors that contribute to infant mortality and better help understand and reduce the burden.

1.3 RESEARCH AIM AND OBJECTIVES

This research project aims to design and develop an analytical platform that will serve as a baseline for an integrated knowledge ecosystem to study infant mortality in the KHDSS. The main results reported in this research project are:

1. Implementation of an interactive knowledge discovery application to study infant mortality.
2. A situational analysis of the current data-ecosystem context of HDSSs and their relation to the platform.
3. Recommendation on how the platform can be used in other HDSS.

1.4 OUTLINE OF RESEARCH REPORT

This research report is organized as follows. Chapter 2 presents the literature review, with relevant works that have been conducted. Chapter 3 presents the methodology and technology employed in the design of the platform. While Chapter 4 presents the use case scenarios of the platform to address objectives 1, 2 and 3. In Chapter 5 , we present the deployment consideration in other HDSS. We finally gave the conclusion and future directions in Chapter 6. Other important details are presented in appendices A, B and C.

2

LITERATURE REVIEW

This chapter presents the review of relevant literature on key concepts on infant mortality, data visualization, and analytics. The chapter reviews related works and techniques on the development of visualization applications for public health research. The reviewed literature is presented as seven subheadings namely: overview of infant mortality, maternal health factors, immunization, child health factors, visualization, visual analytics application, and reproducible research.

2.1 OVERVIEW OF INFANT MORTALITY

Infant mortality is a major global health challenge with 75% (4.5 million) of all under-five deaths occurring during the first year of life [14]. Although global infant mortality has reduced due to the implementation of child survival interventions and social-economic development, the progress is still insufficient, falling short of MDG targets especially in sub-Saharan African and South Asia. With the adoption of the 2030 agenda for SDG by world leaders in 2015, child survival was mapped as one of the health targets aiming to reduce under 5 mortality to at least 25 deaths per 1,000 live births in all countries [15]. This should be achieved by ending preventable deaths in low and middle-income countries as highlighted by Lancet commission [16]. The majority of childhood deaths occur in Africa with the risk of infants dying at 50.13 per 1,000 live births as recorded in 2015 [14]. Although infant mortality in Kenya has gradually reduced from 107.96 deaths per 1000 live births in 1996 to approximately 50.13 deaths in 2015, the causes of death are not reliably attributed [17]. In 2006, Kilifi HDSS data reported an infant mortality rate of 41

deaths per 1,000 live births [6]. The international community led by Bill and Melinda Gates Foundation (BMGF), the Global Fund, governments, and other organizations have increased investment in public health incentives majorly in the fight against AIDS, Tuberculosis, and Malaria. This signifies the importance of health as one of the pillars of the international development landscape for more than 20 years with efforts made to reduce morbidity and mortality universally [16]. Infant mortality is a key population health indicator and has lifelong implications globally. Several studies have examined risk factors associated with infant mortality and have broadly grouped them into: i) maternal health factors; ii) child health factors; iii) demographic; and iv) social-economic characteristics.

2.2 MATERNAL HEALTH FACTORS

It has been indicated that maternal factors i.e. age, parity, birth intervals and utilization of antenatal care services influence child survival [18]. In countries with high mortality, infants born to mothers aged younger than 15 years or old women aged 35 years or older are found to experience more deaths and more so pre-term outcomes. Thus, the impact of maternal age at birth has drawn the attention of many researchers studying risk factors of infant mortality [19, 20]. The reproductive patterns of women and child mortality are well documented with studies showing that mortality rate increases with parity [21] The general thought is that longer birth intervals improves child survival. Closely spaced births or pregnancies have three causal effects namely:

1. Maternal depletion as the mother is not given enough time to recover from adverse physiological and nutritional demands related to pregnancy [22].
2. Sibling competition, therefore subjecting the child to scarce resources including care.
3. Transmission of infectious diseases as well as the severity of infection in a household.

The consequence of closely spaced births are; poor health-seeking behavior, high parity, and breastfeeding for short durations which increases the risk of infant mortality [20].

Available evidence suggests that utilization of Maternal Health Care (MHC) services in Kenya is still low with 58% of women making the recommended number of antenatal care visits of 4 and above during their pregnancy, and 61% of live births delivered in the health facility [23].

Information about MHC indicators displayed through data visualization platforms can help to add meaning to complex datasets and accelerate deepening understanding. Through the use of patterns and trends in outcomes of maternal health highlights the dynamic burden with increased diversity in the magnitude of mortality and morbidity within populations. Visual displays have helped highlight attributes such as maternal health inequality in low- and middle-income countries [24].

2.3 SOCIOECONOMIC DETERMINANTS

Analysis from ongoing demographic and health survey program indicates the existence of huge socio-economic and demographic differences between infant mortality in many developing and developed countries. Parental education and maternal education have a crucial role in the health and survival of a child. Education provides parents with knowledge and attitudes associated with proper childcare, prevention, and management of ill health such as utilizing prenatal care and immunization of children. A study done in Nigeria observed that there was a decline in child mortality as the mother's year of schooling increases [25]. The availability of financial resources in a household means availability of basic human requirements, i.e. place of residence, water and sanitation facilities, which consequently influences infant survival. Infants born in slums and rural areas have a higher risk of death due to lack of adequate resources which increases the risk of exposure to contaminated water, environment and with limited sanitation [26]. Findings from the year 2003 Kenya Demographic and Health Survey (KDHS) showed that the survival of a child in urban areas was determined by maternal level of education and awareness. After

a child has survived the first month of life, ethnicity becomes one of the most important determinants of mortality [27].

With visual representation of social-economic components and its relationship to infant mortality highlights the social-economic disparities among countries. Infant mortality statistics are readily available but with aid of visual analytics, it can lead to less ignorance of its burden and convince countries to major into social-economic development in order to combat diseases and improve population health [28].

2.4 IMMUNIZATION

Childhood immunization guarantees protection against vaccine-preventable diseases and it has been defended as one of the most successful and cost-effective health interventions [29]. One million deaths occur in developing countries every year due to vaccine-preventable diseases and reduced immune function thus increasing susceptibility to infectious and communicable diseases [30]. Countries have different immunization schedules with the majority of the recommended vaccines given within the first year of life. WHO-Global Immunization Vision and Strategy (GIVS) encourages developing nations to vaccinate their infants with multiple vaccine doses and increase their vaccine coverage to at least 90% for herd immunity[1]. GIVS aims to reduce childhood morbidity and mortality to at least two-thirds compared to 2000 levels [29]. From the WHO and United Nations International Children's Emergency Fund (UNICEF) report estimates of immunization, Kenya achieved 80% coverage in the year 2016 [31]. With visual displays and integration to points of data collection, its easier to map out the counties that are underperforming with health being a devolved unit in the government.

1 Herd immunity is a form of immunity that occurs when the vaccination of a significant portion of a population (or herd) provides a measure of protection for individuals who have not developed immunity

2.5 CHILD HEALTH FACTORS

Infants born with low birth weights remain at much higher risk of death than normal-weight infants at birth. Although the percentage of underweight children declined globally as a result of hospital-based services, it remains a problem in sub-Saharan Africa and known to correlate with poor neonatal outcomes in relation to mortality and morbidity [32]. From 2018 KDHS data, the survey reported 8% of children were born with low birth weight (less than 2.5 kg). The percentage of low birth weight reported varied from a low of 4% in the Nyanza region to a high of 13% in the Coast region [23]. Another child-level factor for infant mortality is gender, with studies reporting that males appear to be at higher risk of death than females. This is attributed majorly due to genetic factors and environmental factors leading to adverse reproductive outcomes. The x-linked immunoregulatory genes are associated with greater female resistance to infectious diseases [33].

Mosley and Chen argued in their widely cited 1984 article that child mortality in developing countries is mostly 'the consequence of a series of biological insults' as opposed to a single outcome [34]. The top four causes of death in children under the age of 5 years are infectious diseases predominantly pneumonia, diarrhea, and preterm birth with malaria. Numerous factors may have contributed to the decline of infant mortality in sub-Saharan Africa. Possible factors include: decline of malaria prevalence, changes in immunization rates, giving birth in health facilities and access to safe water and sanitation [35]. Although some studies do not identify these causal effects, other studies observed that maternal health factors have a protective effect on infant mortality [36].

In order to draw accurate interpretations from data collected on mortality, focus has currently been put on the visualization of information [37]. The potential of visual analytics tools in public health investigation that facilitates discussion from detecting anomalies, monitoring trends to generating testable hypothesis has grown. However, increasing pop-

ularity in emerging visual analytics approaches using complex graphs in public health studies are lacking [38].

2.6 VISUALIZATION

The story of John Snow plotting of cholera cases has been elegantly told [39]. Ever since graphs and visualization have played a critical role in epidemiology. Numerous tools and algorithms have been developed to help health professionals in the public health sector to analyze and visualize data thus supporting communication, generation of new hypothesis and making informed decisions [37, 39]. Interest in public health informatics reflects the critical role that data collection, analysis, and utilization plays in health care. Visualization is essential for exploring data and presenting results as it harnesses perceptual capabilities by providing visual insight from data. This includes looking at patterns, trends, features, anomalies and more so the relationships [16]. Data visualization has lagged from its other dependent areas of data science such as data capture, storage, analysis, and knowledge discovery. Besides, there is a capacity gap of the ability of a user to extract answers and present meaningful information. With current advancement in electronic age, disease burden in a population can be mapped in terms of geographical distribution, demographics, and risk factors [40]. Visualization tools help researchers to integrate data, synthesize and visualize information relating to surveillance of disease, prevention, and its control. Geographic Information Systems (GIS) tools have played a critical role in the detection of disease clustering, its spread within territories and even predict outbreaks. Modern approaches to visualization such as phylogenetic trees and social network graphs are increasingly been used to characterize disease outbreaks. Thus, adopting a visual analytics tool can fundamentally improve task performance from predefined algorithms. In today's enterprise solution market, the next breakthrough will come from platforms offering unified solutions for data mining algorithms and visual interface.

2.7 VISUAL ANALYTICS APPLICATIONS

The Centers for Disease Control (CDC), one of the most influential public health organization identified visualization as one of six areas of focus to advance health surveillance in the 21st century. With new approaches to integrating visualization approaches, the analytical process will help adopt the traditional public health functions [41]. New data-driven methods are being scaled up to make visualization the end product of scientific analysis rather than providing exploring capability, thus becoming the central piece of the scientific process [42]. Tableau [43], one of the leading software in visual analytics was founded with a simple mission of helping people understand their data and yield questions that can lead to actionable insights. Tableau was developed for desktop and server environment and is one of the most popular commercial enterprise application software used by various organizations. Qlikview, another major player and the biggest competitor of Tableau, provides an easy way of creating insightful data for visualization [44]. The analytics platform has advanced tools that allow users to interact with their data in a visual environment [45].

With rapid progress in research, a variety of computer programming languages are currently in use with diverse strengths for analyzing and visualizing complex data. R is an open-source statistical and programming environment that incorporates all statistical tests, models as well as analysis with outstanding graphical capabilities. R has a web API called Shiny that generates interactive data visualization on the dashboard [43, 46]. Shiny web development allows customization for programmers to show R out-puts displayed interactively on a web browser. Moreover, Rmarkdown and knitr package for R and Git for version control are essential for data analysis and producing dynamic documents. Metabase also an open-source application, enables the development of data analytics and business intelligence applications or platforms. Metabase generates GLOPR descriptors

for object persistency in an automatic way and manages inheritance among meta model classes [47].

The concept of visual-analytics applications started with open-source tool kits in academic research institutions but in the recent years we have observed big commercial software vendors such as IBM, Microsoft, and SAP integrating VA components into their frameworks that would eventually promote studying of public health data and ultimately enhance decline in all forms of childhood mortality [48]. Therefore, developing new tools with intuitive features that integrate data aggregation to visual analytics and reproducibility of result amid little training would be essential [40].

2.8 REPRODUCIBLE RESEARCH

Replication of scientific findings is a core principle of science by which claims are reproduced. The more frequent a given relationship is observed by independent scientists, the more we build trust and credibility of published scientific research and accelerate discovery. Increased data generation comes with the danger of identifying spurious associations leading to misleading findings. In addition, investigations cannot be fully replicated due to lack of resources and time in many fields of scientific study [49]. To fill the void between full replication and evaluating nothing, reproducible research is the alternative of verifying published results from data sets and methods provided and conducting alternative analysis. Epidemiological studies are the cornerstone of public health and are instrumental in identifying risk factors associated with disease and are critical in setting up policy for decisions. Thus, making replication of great importance where findings are substantiated for preventive healthcare. The prerequisite for reproducibility is good for population health data which has influenced the population health approach in terms health outcomes and its distribution and goes beyond focusing on the health status of an individual [50]. Current epidemiological studies are with relative risks and are more

easy confounded or with reporting biases. Replication of results from studies will provide proper interpretation of results and potentially reduce infant mortality [51].

INDEPTH facilitates longitudinal population-based surveillance for researchers from HDSS sites in low and middle-income countries(LMICs). Major feature of INDEPTH include multi-center initiatives such as comparative analysis with cause-specific mortality patterns and effects of changes in climatic conditions [52]. To enhance research data management capacity, INDEPTH also conducts training for data managers through iSHARE (INDEPTH Sharing and Access Repository) to affiliated HDSS member sites by standardizing events files (births, deaths, and migrations) from surveillance data. The data is shared globally through an online repository which is updated annually [53]. INDEPTH also provides a platform for visualization of key demographic indicators using INDEPTHstats website from data held in the repository. The website assists users to make a quick overview of the information per site which has been used for quantifying MDG, provide cause-specific mortality and describe the impact of major infectious diseases and relevant interventions [54]. KHDSS has been collecting infant mortality data for more than 10 years. This has aided the evaluation of trends that lead to assessment of MDG 4[2]

2.9 SUMMARY OF THE LITERATURE

A review of the literature dealing with risk factors for infant mortality was conducted and various studies reviewed. However, there was no focus on the utilization of automated visualization platforms globally for studying infant mortality. This underscores the need for the development of an automated visualization application to answer questions on child health and surface insights that create impact.

This research on integrated knowledge ecosystem platform at HDSS site will promote the implementation of analytical data visualization tools and technologies allowing it to become integral for the scientific process in public health.

2 Reduce child mortality

3

DATA MODEL FOR THE INTEGRATED KNOWLEDGE ECOSYSTEM

3.1 INTRODUCTION

This chapter focuses on the methodology used in the study by describing the core KHDSS data and its integration with other studies. Also presented in this chapter are details of variables used as well as data storage systems and data analysis procedures.

3.2 DATA SOURCES

The study utilized 2015 Kilifi HDSS dataset of children aged less than one year from the study population. KHDSS is one of the 52 HDSS sites affiliated to INDEPTH network. These data was obtained after getting approval from data governance committee at KEMRI-Wellcome Trust and ethics clearance from the University of Witwatersrand. The core KHDSS data collected at each re-enumeration round after the baseline enumeration include;

- Births
- Pregnancy
- Migrations
- Observation (Events)
- Deaths
- Homestead

The demographic surveillance data is stored in an Integrated Data Management System KIDMS which runs on a MySQL service. Data from other studies like the Vaccine Monitoring System (VMS) are also nested on KIDMS and linked through a unique identifier.

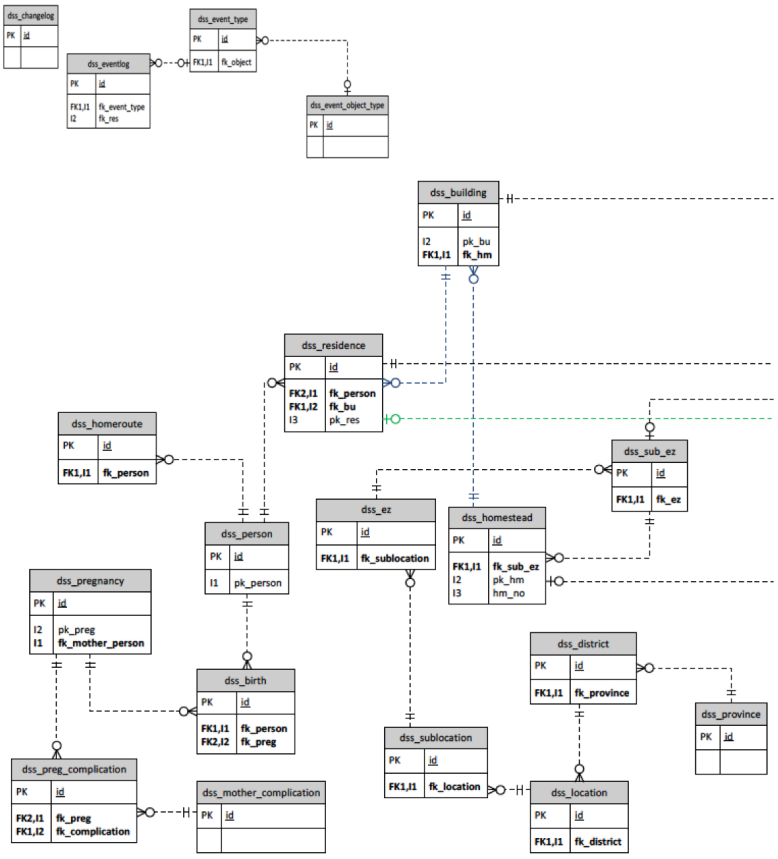


Figure 1: Kilifi HDSS Data Model

3.3 KILIFI HDSS DATA MODEL

To monitor population changes that are dynamic, HDSS sites use a Relational Database Management System (RDBMS) with each site having different data models. The implementation in each site is built on different proprietary and open-source RDBMS system such as MySQL, Microsoft FoxPro, PostgreSQL and Microsoft SQL, etc. The INDEPTH entity relationship diagram (see Figure 1) enables the sharing of data between sites. The tables in the database are linked together to form the relational structure in terms of relationship between individuals who were enumerated in the surveillance area.

The Kilifi HDSS data model differs slightly with the core INDEPTH model in that, Kilifi stores migration data on the events log table and missing relationship status to the head of the household.

3.4 DESCRIPTION OF CORE DATA

The following are the descriptors of the core KHDSS core data:

- **Births:** A person born to a resident mother and forms part of new entrant into longitudinal cohort. Births are significant for the analysis of fertility and the generation of infant mortality or neonatal deaths.
- **Migrations:** Migration involves tracking movement of people into/out of surveillance area, usually within the study area or outside the study area. Operational definition differs from each HDSS site and usually involves recording additional information like reason for migration and the destination. Migrations form part of new entrant or exit to the cohort.
- **Deaths:** Death forms part of the exit to the cohort. Event recorded with documentation by age, sex, with the underlying course of death recorded in verbal autopsy form. Deaths are critical in the calculation of mortality rates within the surveillance area.
- **Pregnancy:** Pregnancy information is collected from all resident women who are within the reproductive age (15-49) not limited to. Information collected includes: LMP, whether pregnant and attendance to ANC
- **Observation:** Observation involves recording all type of events associated with each individual, the date the event was observed and when the event actually occurred.
- **Homestead:** A homestead is the dwelling place in which a resident is located. Information collected includes: the physical structure and the corresponding geo coordinates.

3.5 VARIABLES AND VARIABLE MEASUREMENTS

The dependent and independent variables for this study were selected at one time-point i.e year 2015 from KIDMS surveillance database and Vaccine monitoring system (VMS).

3.5.1 *Dependent Variables*

The dependent variable in this work was infant mortality (defined as death before the first birthday)

3.5.2 *Independent Variables*

The independent variables in this work included: Infant characteristics at the individual level, infant immunization information and maternal information. The selection of these independent variables was guided by reviewed literature and Data Governance Committee (DGC) data approval. Table **??** presents the definition of independent variables.

3.6 STATISTICAL ANALYSIS

Statistical analyses were performed using the R software/environment. Descriptive statistics were presented in terms of mean, median, maximum and minimum for continuous variables. To assess for the risk factors of infant mortality, we used the logistic regression model to analyze data on over 5000 infants.

Table 1: Definition of variables

Variables	**Definition**	**Coding**
Sex	Sex of the child	Male (1) Female (2)
birth locn	Location of birth	Homestead (1) Traveling (2) Hospital (3) Hospital
death locn	Location of death	Homestead (1) Traveling (2) Hospital (3) Other (4)
Longitude	Distance north or south of the equator	Recorded in degrees and minutes
Longitude	Distance east or west on the earth's surface	Recorded in degrees and minutes
Parity	No of living children child's mother had	(1) <3 children (2) 3-4 children (3) 5+
ANC	child's mother attended ANC	No (0) Yes (1) Don't know (99)
preg outcome	pregnancy outcome	preterm delivery (0) term delivery (1) miscarriage (2)
gestation	duration of pregnancy	Numeric response in weeks
age	Age of the mother at birth of the child	Numeric response in years
vac status	vaccination status for the child	Fully (1)partially (2) not vaccinated (3)

3.7 OVERVIEW OF THE TECHNOLOGY USED

The application was developed using the R programming language and utilizes Shiny code embedded on R Markdown and hosted on RStudio server. Shiny R package provides a framework for developing interactive web pages while R Markdown assists in creating dynamic documents.

3.8 R (PROGRAMMING LANGUAGE)

R is an open source programming language for statistical computing and data science. The capabilities provided by R are extended through user-submitted packages for certain functions with the core packages integrated with installation of R. We preferred R primary because of the software ability to handle data manipulation, analysis, graphical display, integrate with shiny and creation of dynamic documents.

3.9 R SHINY

Shiny is a web application framework for building interactive web applications with R as the backend programming platform to transform analysis into a visual dashboard. Shiny applications can be extended with HTML widgets, Cascading Style Sheets (CSS) themes, and JavaScript action.

3.10 R MARKDOWN

R Markdown is an authoring framework rendering file format for making dynamic documents[1] with R. R Markdown supports a variety of file formats including HTML, MS Word, PDF, HTML5 slides, books, websites, scientific articles, shiny applications and many more which can either be static or dynamic. R Markdown documents are fully reproducible and use multiple languages including Python, R, and SQL. Shiny code can be embedded in a chunk of R Markdown code to make the document more interactive.

3.11 DEPENDENCIES

The platform is dependent on several R packages including; ggplot2 for elegant visualization, maptools for manipulating and reading shapefiles, dplyr for efficiently manipulating datasets, plotly for analyzing data into grid and allowing graphs to be embedded or downloaded.

3.12 ARCHITECTURE

The main components of the developed platform involved the R shiny UI (user interface) object and a server script. The two components were passed as arguments to the shinyApp (embedded on R Markdown) and were combined with additional scripts, data and other resources for the support of the application. We deployed the application on R studio server installed on CentOS version 6 (Linux distribution). The R studio server allows anyone with internet connection and authentication to use the web platform. The system explores and transforms data into interactive dashboard as means of analyzing

1 A document or a web page that includes content taken from external sources that changes as those external sources change. Its content is usually based on the user preference.

and understanding data with different visualization layouts using pre-defined algorithm as highlighted in Figure 2 and in the process below:

- Source data is uploaded directly to the platform as a flat file(.csv) and used to construct data tables for internal representations.
- The resulting data tables are then subjected to R graphical libraries ggplot2 and plotly for visual mappings to generate a visual abstraction and DT package for table representation. The resulting data structure introduces new visual features creating spatial layout, shape with predefined, size and color.
- The actual rendering of data deployed by functions of R is done at this stage into a number of interactive views.
- User interaction with the visualization platform is invoked at this stage. Manipulation through control widgets help sending messages to the application such as fileInput function, selectInput function, sliders and drop down list.
- Shiny package facilitates connection as it 'speaks' to web sockets.
- The R studio server receives requests from end-users then transfers to the R engine for processing through the web browser. HTML widgets are included in the platform and assign a render call to the output of the server side e.g. renderPlot and tableOutput

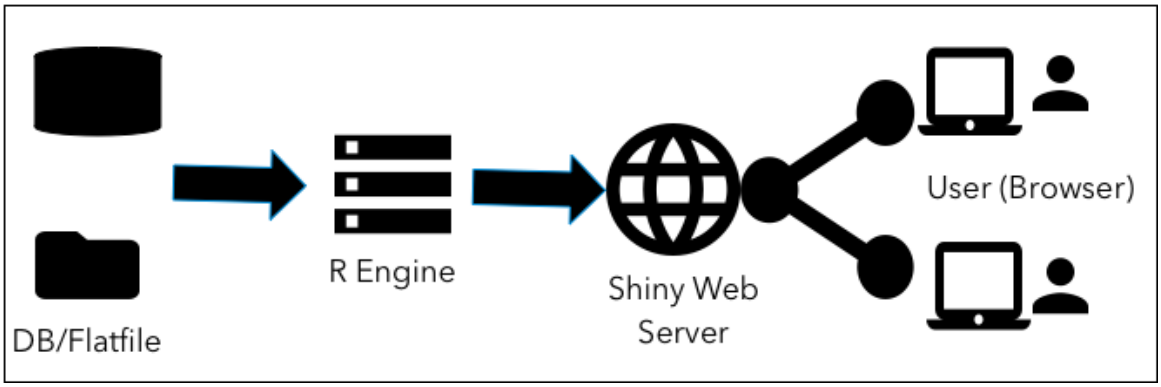


Figure 2: R Shiny Architecture

3.13 MODEL

The model included setting up of user interface (UI) and the server-side for handling data for pre-analysis and visualization. The user interface involved:

- Predefined layout functions.
- Adding elements.
- Layout the app using grid the system.
- Using themes and CSS for styling.
- Adding widgets.

Server interface included:

- Listen for changes with reactive or observe objects
- Link of user interface with server by render*or Output* to display tables, text, and plots.
- Adding dynamic UI

Shiny widgets which is an interactive element on the user interface allows users to explore the data input. The parameters are called by the Server.R and processed by the reactive expression for output or the render* functions and sent for display on the user interface (UI). Installed Shiny package comes with pre-build R functions and logical strings of arguments. Widgets ensure users only select plausible value, the application has widgets for selecting the input variable or groups for the dependent and independent variables.

Shiny *Output functions which are defined in server.R call the reactive R objects. * denotes object such as table (for display of data view and summary statistics), plots (for graphs and maps) as well as text (for regression interpretation of output)

The server controls the user input which then gets displayed on the user interface. Once the application is initiated by data input, loading of package libraries, themes and functions will be executed. The dataset and infographics (with static display) tab will display the uploaded KHDSS data. The other remaining tabs (with dynamic representation) i.e graphs, exploratory analysis, regression, trends, and spatial distribution which have reactive expression (render* function) will be executed depending on the change of a widget. E.g For logistic regression, after choosing any predictor and response variables the user displays the output (statistical significance and effect size) by clicking the reactive action button for the output object(table and text). The other remaining tabs, plotting of outcome depend on re-execute of every reaction of widget change.

Features of the platform included: Data inspection, cross-tabulation, Univariable/multivariable logistic regression, trends display and spatial analysis. All the results from different tabs are downloaded as .csv or .png for plots. The platform allows navigation and zooming options. Since the application is hosted on a Shiny Server, users can concurrently access the application.

The Shiny package of R helped to visualize and performed automated analysis by studying infant mortality from the KHDSSdata and demonstrated data visualization as a promising alternative for display and communication of public health data.

3.14 SUMMARY

The outcome of the data visualization platform will foster improved hypothesis generation, especially in the quantitative skills gap. Furthermore, data visualization can greatly improve public health's ability to tell stories. We present several use cases in the next chapter as proof of concept and describe the platform's components.

4

UTILITY TOOLS AND MODULES FOR THE INTEGRATED KNOWLEDGE ECOSYSTEM

This chapter presents the features of the platform. Thereafter discuss the use case of the application and results generated using dataset from Kilifi HDSS.

4.1 INTRODUCTION

We illustrate the performance of the application by using a dataset of 5,288 records. There are six utilities from the design layout implemented at each tab representing an analytic entity within the platform. Each entity or module provides an easy to use graphical user interface focusing on analyzing infant mortality data. In this chapter, we discuss these modules by demonstrating the features available and input or output data. To carry out these pre-analytics, users need to upload data files, then process the output through the generation of tables of values and plots with results.

4.2 DATASET TAB

The dataset tab consists of three panels (see Figure 3) with a data upload section. The panels are dedicated to help the user understand the corresponding uploaded dataset.

these are:

1. **Data view:** Each column on this panel represents a variable while each row represents a case. The view allows filtering by column i.e. searching a word in any of the

columns or the entire table. It is also possible to sort the rows in an ascending or descending order.

2. **Data structure:** Displays the number of variables, number of observation and the data class.

3. **Summary statistics:** Provides a quick and simple description of continuous variables. Includes mean, median, mode, the minimum value, maximum value, range, standard deviation, skewness, kurtosis and standard error.

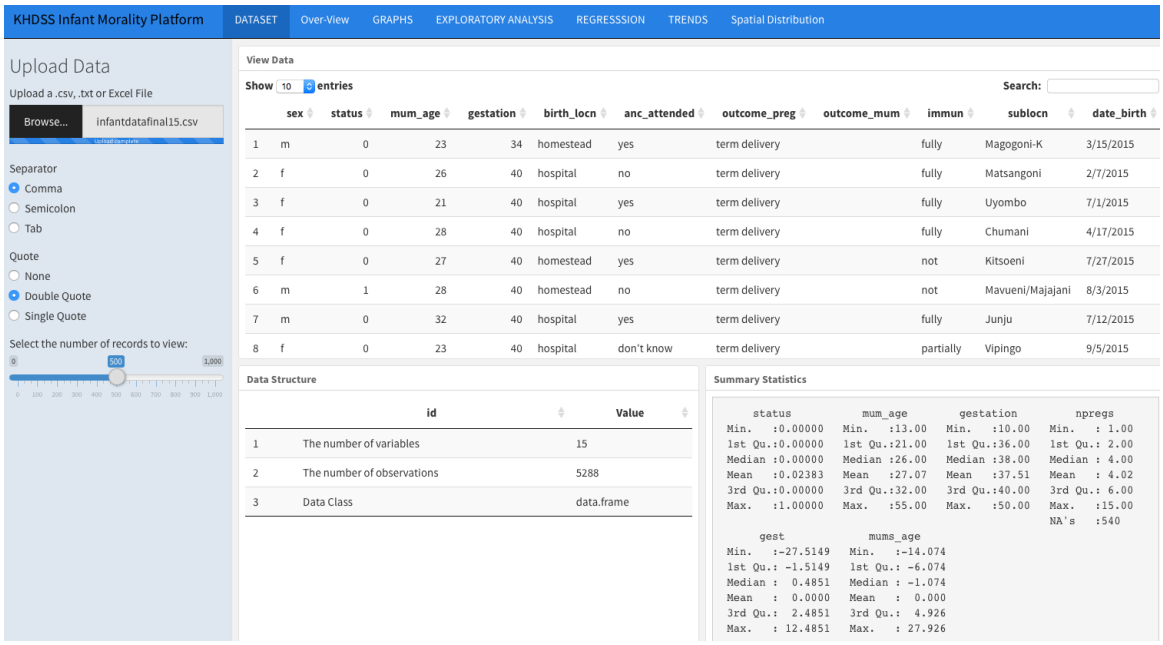


Figure 3: Screenshot showing data upload sidebar with dataset variables, data structure, and summary statistics panels

The dataset tab displays part of the descriptive statistics of the uploaded data and the sample size. Summary statistics for maternal age at birth, parity and gestation period which were part of the main factors influencing infant mortality are displayed for researchers. Figure 3 displays mean of 27 years, 37 months and 4 for maternal age, gestation period and number of previous pregnancies respectively. These figures are usefully in aiding policymakers in decision making e.g Informing mothers the importance of family planning as a pillar of improved child health outcomes with high numbers of previous pregnancies observed.

4.3 INFOGRAPHICS TAB

The infographics provides an overview of the KHDSS data with key information about the uploaded data (See Figure 4). The tab displays the total number of observations, female-male ratio, ANC coverage, immunization proportion, infants mothers age groups and the KHDSS map with infants' death rates.

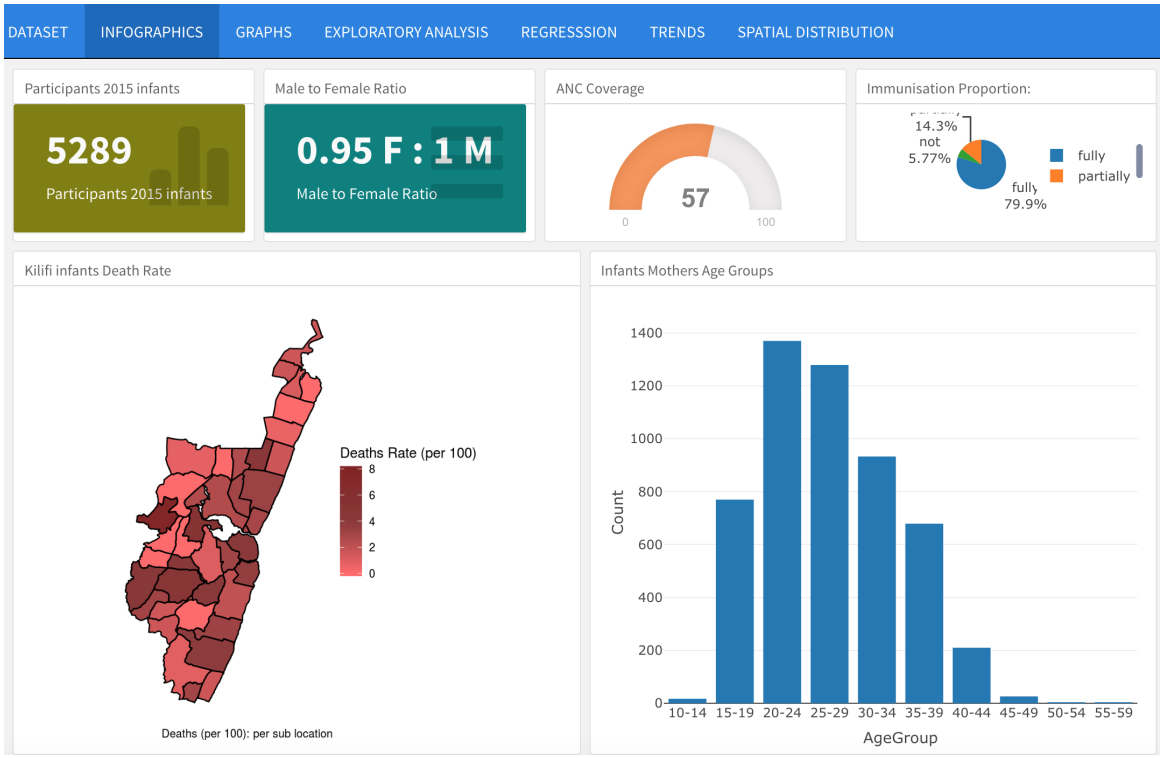


Figure 4: Screenshot showing KHDSS infographics

GIS tools have helped public health researchers to detect disease clustering. Figure 4 displays infant mortality spatial data at KHDSS, locations with dense color indicates areas with more infant mortality rate. The Antenatal Care (ANC) coverage is wanting with only 57% of pregnant women attending ANC. The data can generate smart hypothesis to ascertain if mortality risks maybe be as a result of differences in health-seeking behavior.

4.4 GRAPHICS TAB

The graphics tabs assist the user to visualize and summarize aspects of the dataset uploaded by providing familiarity with the data. After selection of a particular variable, the system automatically uploads specific graphs e.g if the input variable is continuous and group variable is categorical then platform displays a box plot while if all the input variables are categorical the platform uploads bar graph (See Figures 5 and 6)

The box plot displays can aid policymakers verify the mean age of mothers with reference to place of birth which is also a determinant of infant mortality. Figures 5 indicates that women with a lower mean age (26 years) give birth at hospital compared those women who give birth at home (28 years). From the data, policymakers need to put more emphasis on older women on the importance of giving birth at the hospital.

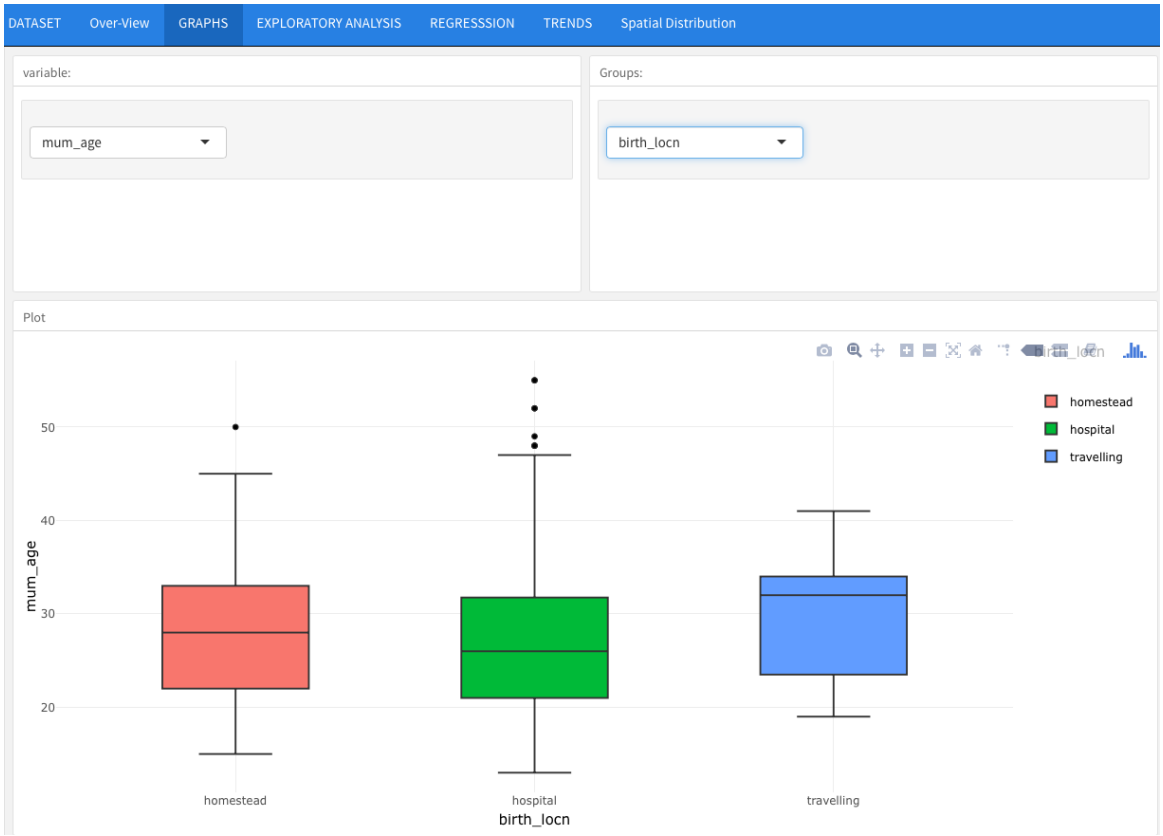


Figure 5: Graphics tab display: A box plot of mother's age stratified by birth location

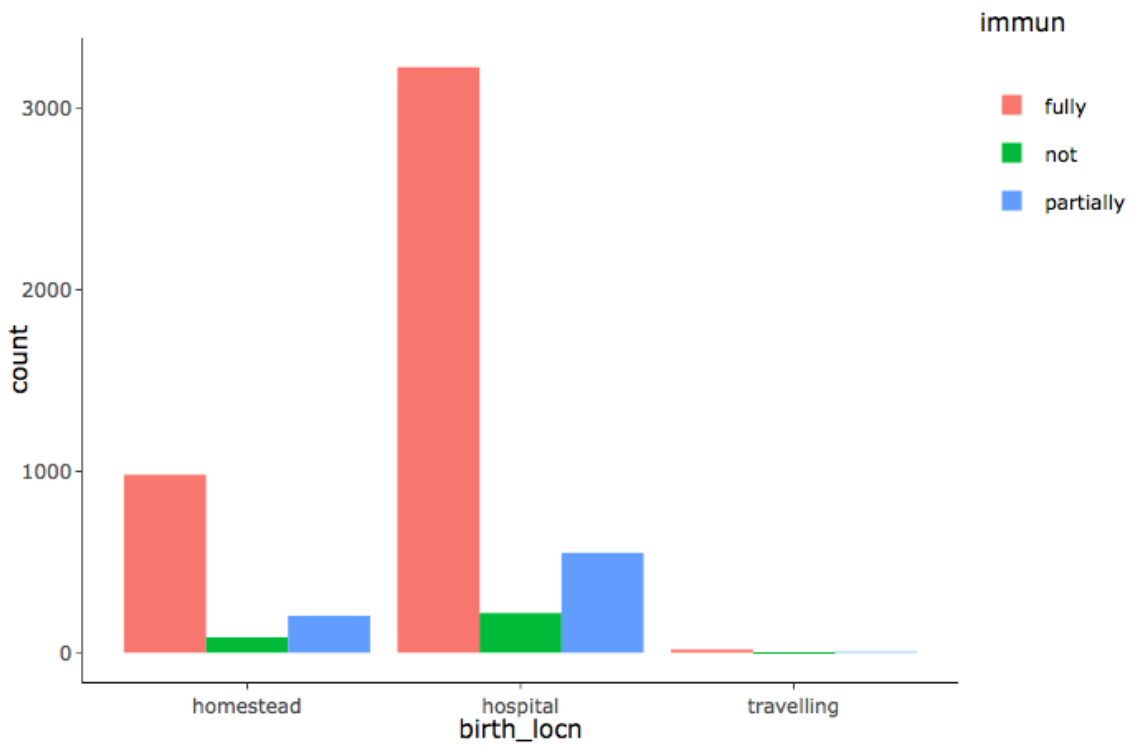


Figure 6: Downloaded graph from the graphics tab display: A bar graph of immunization status stratified by birth location

4.5 EXPLORATORY ANALYSIS

The exploratory tab allows one to perform tabulation of data in rows and columns. The platform can perform either one-way tabulation (data tabulated to one characteristic) or tow-way tabulation (data tabulated on two characteristics). Tabulation summarises data readying for further analysis as shown in Figure 7.

Descriptive results from the cross-tabulation further showed that the majority of infants who died were not vaccinated (23.6%) compared to 5.7% and 0.2% for those partially and fully vaccinated respectively. It is recommended that strategies to address poor vaccine coverage to aid the reduction of infant mortality be strengthened.

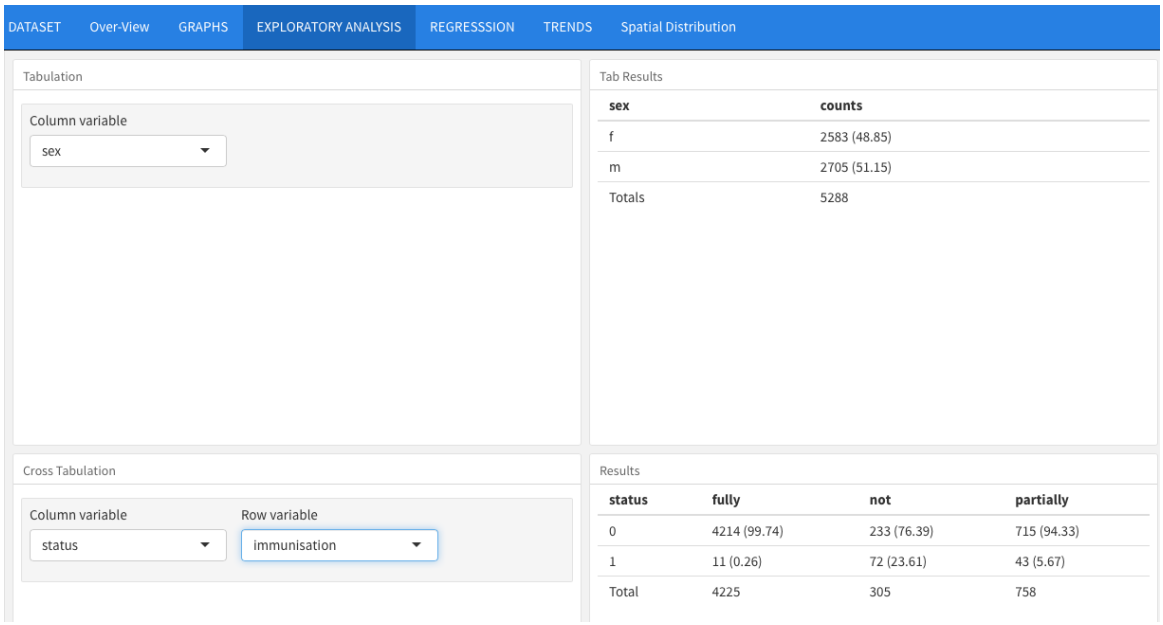


Figure 7: Screenshot showing tabulation tab

4.6 REGRESSION TAB

This section allows the user to load-dependent and independent variables then launch the analytical process by a single click (See Figure 8). The regression tab aims to simplify the task of establishing association and highlights variables that appear to have a significant relationship for the user. The Generalized linear model (GLM) function was fitted with the R packages DT tables for table representation.

4.7 TRENDS TAB

Monthly Infant mortality rates per 1000 live births for the year 2015 are presented in Figure 9. The month of August 2015 recorded the highest death rate at 35.5 while month of January recorded the lowest death rate.

We recommend a longer duration of data instead of one year for the study of infant mortality trends to assist in policy recommendations based on the characteristics of the data.

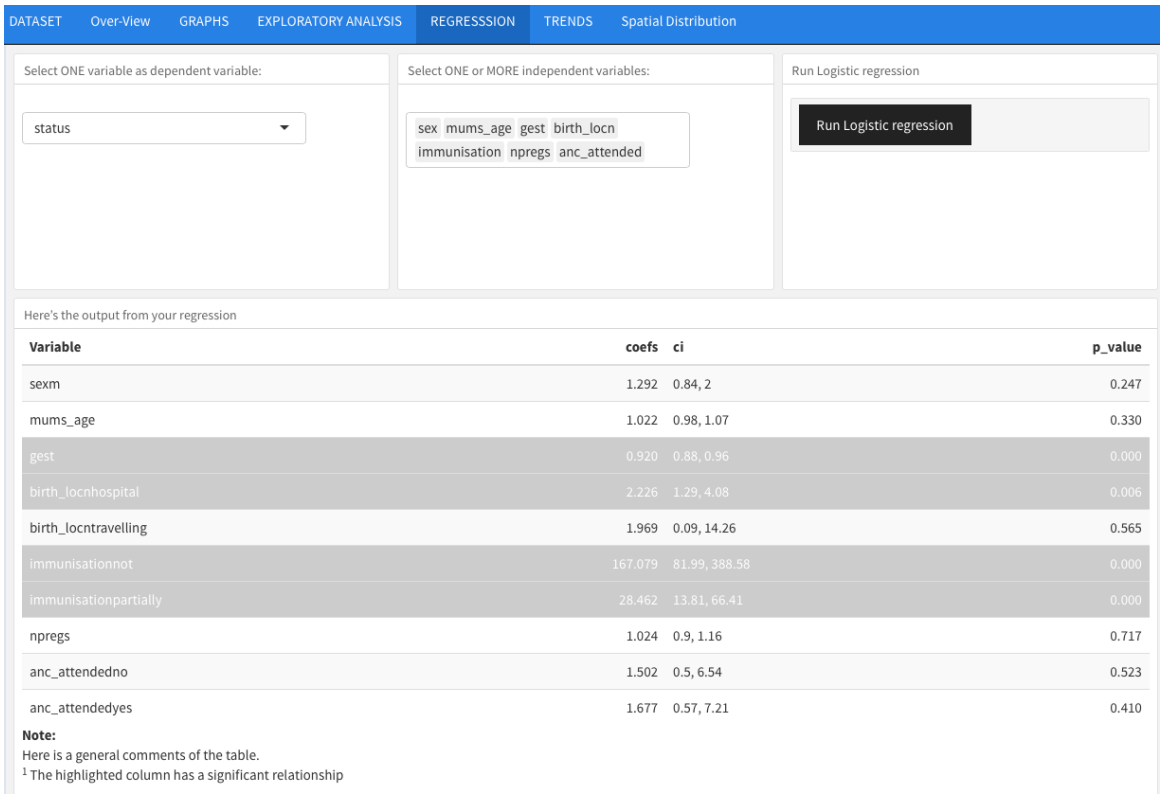


Figure 8: Regression tab: Logistic regression display



Figure 9: Trends Tab: KHDSS infants Death rates(2015)

4.8 SPATIAL DISTRIBUTION TAB

We illustrated the spatial distribution of infant deaths across the Kilifi HDSS surveillance area for the year 2015. First, we needed to upload the data file with coordinates (.csv), then we uploaded the shapefiles containing the areas of study region (Kilifi KHDSS study area). After uploading all the required files, the application automatically checks if the files are correct then loads the results in the output section (See Figure 10). From the map, most deaths are concentrated within Kilifi township and sparsely distributed within the rural areas. Scaling of intervention in areas with more infant deaths in the community will play a significant role in reducing infant mortality and also set up studies for causal relationship.

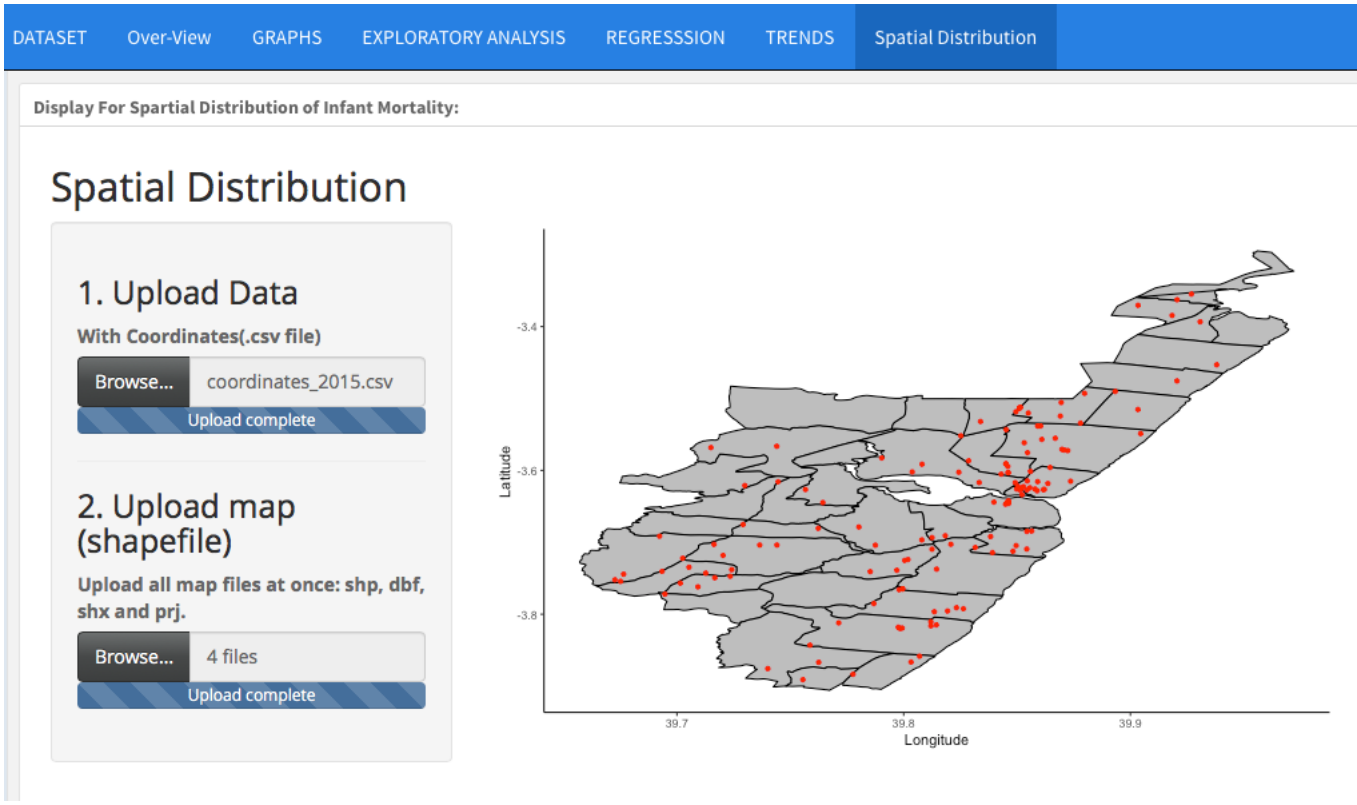


Figure 10: Spatial distribution of infant deaths in KHDSS

4.9 RESULTS

Table 2 gives the descriptive statistics of variables stratified by survival status (either dead or alive) for data obtained in 2015. Of the 5,288 infants, 2,705 (51%) were females while 2,283 (49%) were males. The infant's mother's age ranged from 13 to 55 years, with mean a age of 27.0 and 26.5 for those infants who were alive and died respectively. The vaccination status varied widely among infants who died with almost a quarter 72 (23.6%) reporting to have been unvaccinated while 43 (5.67%) were partially vaccinated and only 11 (0.26%) reported to have been fully vaccinated. With respect to birth location, 106 (2.65%) infants were born in hospital died while only 19 (1.5%) infants and 1 (4.17%) infant born at homestead and while traveling died, respectively.

Table 2: Descriptive Characteristics

Variable	**Category**	**Alive n(%)**	**Dead n(%)**
Mothers Age (mean,SD)		27.09(6.96)	26.50(6.92)
Gestation (mean, SD)		37.56(4.24)	35.67(2.38)
Previous Pregnancy (mean,SD)		4.02(2.44)	3.89(2.38)
Sex	Male	2632(97.3)	73(2.7)
	Female	2530(97.95)	53(2.05)
Immunization Status	Fully vaccinated	4214(99.74)	11(0.26)
	Partially Vaccinated	715(94.33)	43(5.67)
	Unvaccinated	233(76.39)	72 (23.61)
ANC Attendance	Attended ANC	2957(97.62)	72(2.38)
	Not Attended	1858(97.48)	48 (2.52)
	Don't Know	347(98.3)	6 (1.7)
Birth Location	Homestead	1250(98.5)	19 (1.5)
	Hospital	3889(97.35)	106 (2.65)
	Travelling	23 (95.83	1 (4.17)

Table 3: Logistic regression analysis results

Variable	**Category**	**OR(CI 95%)**	**P-Value**
Mothers Age		1.02 (0.98, 1.07)	0.330
Gestation		0.92 (0.88, 0.96)	<0.001
Number of Pregnancies		1.02 (0.9, 1.16)	0.717
Sex	Female	Ref	
	Male	1.29 (0.84, 0.24)	0.247
Immunization Status	Fully vaccinated	Ref	
	Unvaccinated	167.08(81.99, 388.58)	<0.001
	Partially Vaccinated	28.46(13.81, 66.41)	
ANC Attendance	Don't Know	Ref	
	Not Attending	1.50 (0.5, 6.54)	0.665
	Attending	1.68 (0.57, 7.21)	
Birth Location	Homestead	Ref	
	Hospital	2.23(1.29, 4.08)	0.024
	Travelling	1.97(0.09, 14.26)	

In the logistic regression analysis[1], gestation period, birth location and vaccination status were significant predictors of mortality. We found that gestation period was associated with a 8% reduction in infant mortality (OR = 0.920, 95% CI: 0.98, 1.07 (p<0.001)). On the contrary, infants born in hospital were 2 times more likely to die than infants born at the homestead (OR = 2.23, 95% CI: 1.29, 4.08 (p=0.024)). Mortality was significantly higher among the unvaccinated infants (OR = 1.969, 95% CI: 81.99, 388.58 (p<0.001)) and partially vaccinated (OR = 28.462, 95% CI: 13.81, 66.41 (p<0.001)) compared to infants who were fully vaccinated. Mortality was not associated with sex, mother's age or attendance to antenatal clinic. See Table 3 for logistic regression analysis result extracted from the platform.

1 Function to model binary dependent variable

4.10 SUMMARY

The findings generated by the automated visual analytic platform indicate that infant mortality remains a public health challenge. The application further displays infant mortality variations per location based on GIS data which can relate to the national level trends in Kenya.

The platform proved to be an interesting and valuable analytical tool to health care workers as time spend on analytical work is expedited. Measures aimed at addressing issues of infant mortality can be pursed based on the results generated by the application.

With the advocacy of open access, the platforms code is immediately available without access restrictions that allow other users to read, download, and with concomitant collaborative development for the application improvement or introducing new functionality. Therefore, the developed application is carefully documented for deployment purposes as well as to attract visibility and contribution from the community.

5

USING THE INTEGRATED KNOWLEDGE ECOSYSTEM IN OTHER HDSS

This chapter presents the deployment consideration of the platform to other HDSS sites, the prerequisites, and the infrastructure consideration.

5.1 A SITUATIONAL ANALYSIS OF THE CURRENT DATA-ECOSYSTEM CONTEXT OF HDSSS AND THEIR RELATION TO THE PLATFORM

The effective design of HDSS sites databases are essential and provides the potential for synergy and pooling of data from multiple studies that maintain the dynamic population. The data usually has three main dimensions;

1. **The entity**: Relates to the set of items been described which should be captured and stored in the form of properties, tables or workflows.
2. **Its attributes**: Refers to the database component that describes each entity.
3. **Time**: Describes the ever-evolving relationship between entities.

HDSS sites have each customized their site-specific databases around a Reference Demographic Surveillance Data Model (RDM) [55] as shown in Figure 11. These customizations include changes in the database schema that are built on different proprietary RDBMS and laid on different generic system for data capture as well as 'Open HDS' as shown in Table 4.

The structure of RDM is highly normalized therefore reducing the likelihood of data inconsistency. In order to respond to a wider spectrum of HDSS needs, the RDM allows

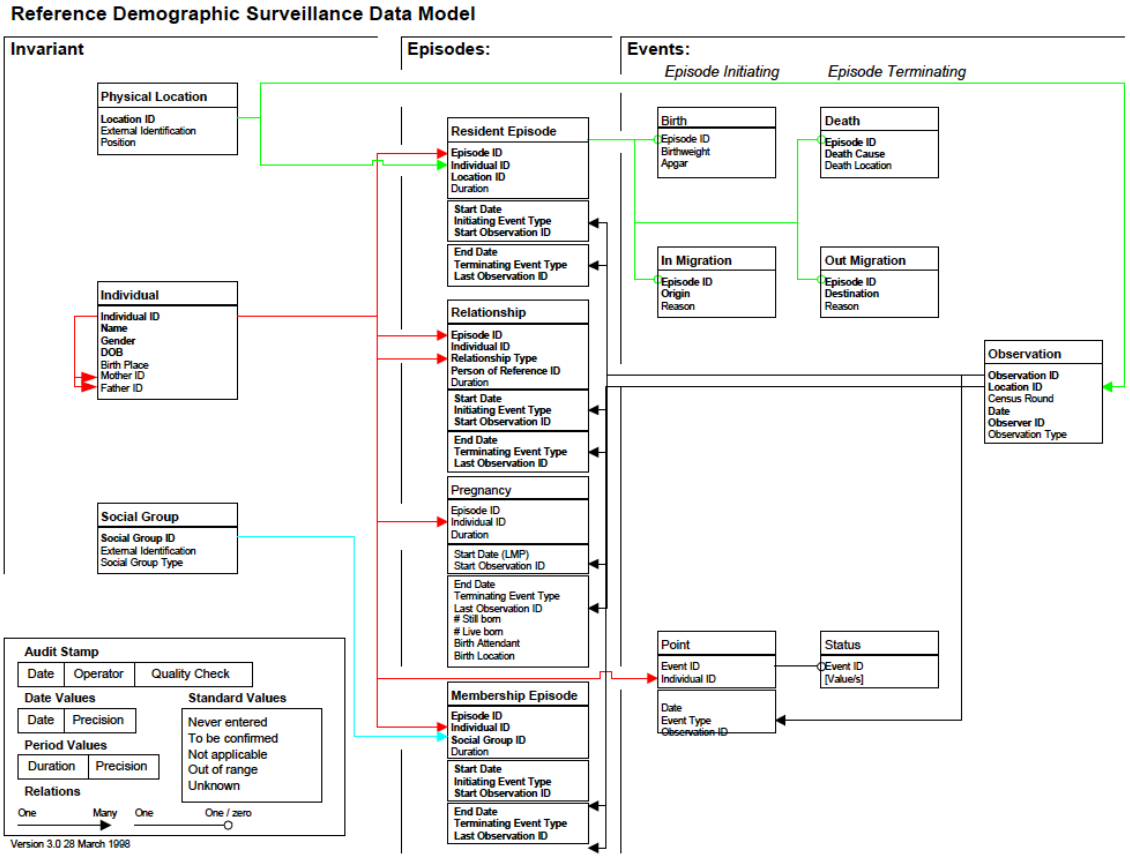


Figure 11: Reference Demographic Surveillance Data Model

easy integration of additional components depending on the site needs without losing its clear overall structure.

Kisumu, Harar, Kombewa and Agincourt HDSS sites adopted the RDM entirely. However, Kilifi and Moramanga HDSS have notable difference compared to the four sites which include:

- With the Kilifi and Moramanga HDSS database schema, all events (i.e. births, deaths, and migrations) are logged as a single entity and categorized by the type of event. While in Kisumu, Harar, Kombewa, and Agincourt each event is logged as a single entity. The major drawback of utilizing a single entity for all events is the potential of having a huge volume of data in the events table that might slow down the application or the database.

Table 4: Sample implementation at 6 HDSS sites

HDSS Site (Country)	RDBMS	Application	Data Collection tool
Kilifi (Kenya)	MySQL	Customized web-based PHP	Electronic (Windows Tablets)
Kisumu (Kenya)	MS SQL	Customized VB.Net	Electronic (Netbooks)
Kombewa (Kenya)	MS SQL	FoxPro	Electronic (Netbooks)
Harar (Ethiopia)	MySQL	Open HDS	Android Tablets 3rd party web server
Agincourt (South Africa)	MS SQL	Customized	Redcap
Moramanga (Madagascar)	MySQL	Customized web-based PHP	Android Tablets 3rd party web server

- The other difference is that the RDM has a social group table that captures 'a full family tree' with each individual associated with one or more social group, for Kilifi HDSS there is only self-referencing on mother to child.

- The other remaining entities are similar with differences in how tables and variables are named. e.g. tables are either named individual or person, resident episode or residence, event or events log.

- In some sites (Kisumu and Moramanga) the assignment of the unique person identifier (PID) utilizes a composite key i.e. they construct their PID by combining keys from different tables by use of location, village, household and member-number (uniqueness is guaranteed when columns are combined) while in Kilifi and Agincourt an auto-increment keys is used.

- Apart from the differences in the database schema structure and the type of application for data collection, HDSS sites have operational differences in ways of data collection and definition of terms e.g. in Kisumu dwelling(s) are defined as a households while Kilifi dwelling(s) are defined by homesteads with no reference to those who eat from the same cooking pot like the other HDSS sites.

- Each year, almost all HDSS sites have a special cross-sectional census module nested within the surveillance system to collect social-economic data (SES) and additionally bed net usage data for Kilifi HDSS.

5.2 DEPLOYMENT CONSIDERATION AND DATA LINKAGE

The key strength of HDSS sites is the integration with other research studies. To create a single dataset in a standard format from the core dataset [21] and other platforms like cohort studies, observational studies, clinical trials or national surveys, can be a daunting task but achievable. The INDEPTH Data Management Programme (IDMP) was formed to enhance research data management capacity to member centers and enable them to perform data extraction from different data sources, pre-processing then harmonizing and creating an analytical dataset [53]. For studies missing automated linkage to the core HDSS data will utilize the Extract Transform and Load (ETL) process by use of business intelligent suites like Pentaho data integration program or by use of SQL queries and configuring the data into standard variable format see figure 12 for data upload. The ETL capability makes the data more responsive at run-time and available in a flat-file format.

The method for deploying the integrated knowledge ecosystem application varies with respect to the type of data model and, the organization's infrastructure. Whatever the outcome of interest, the developed application can be used to measure its effect such as mortality, morbidity, assessment of effectiveness of bed net, factors for malnutrition etc. The INDEPTH Network member HDSS sites use different database structure to store their corresponding surveillance data. However, all participating sites have a common event attributes for the core dataset as shown in Figure 12 and a range of different research studies as from the sample listed in Table 5 [6, 56–58].

Based on reports published by Kisumu and Dodowa HDSS with focus on data linkage and integration [56, 57] with a strategy of deploying the application. The linkage process at points of health contacts relies on matching centered on self-reported characteristics using basic demographic information (Names, date of birth, Location and Landmarks) to determine if the said individual already exists in the population register thus avoid multiple registration and thereafter assigned a unique identifier in the database see figure 13.

Attribute	Variable name	Description
Individual identifier	IndividualId	A number uniquely identifying all the records belonging to a specific individual in the data file.
Location identifier	LocationId	Unique identifier associated with a residential unit within the site and is the location where the individual was or became resident when the event occurred.
Date of birth	DoB	The date of birth of the individual
Event	EventCode	A code identifying the type of event that has occurred for (Birth, Enumeration, In-Migration, Out-migration, Location exit, Location entry, Death, Delivery, Observation)
Event date	EventDate	The date on which the event occurred
Observation date	ObservationDate	Date on which the event was observed (recorded), also known as surveillance visit date

Figure 12: Core data: Common event attributes for the INDEPTH data specification

Kisumu HDSS additionally uses biometric finger scanners to uniquely identify individuals while Dodowa HDSS assigns unique permanent identifier number and issues a card to individuals within the surveillance area. The two sites supplement(validate) names and date of birth, not replacing them [56, 57].

For HDSS sites with different data structure and not affiliated to INDEPTH, users have to create a flat file (.csv format) with a binary dependent variable (two categories labeled as “0” and “1”) then add other desired risk factors as independent variables either as continuous, discrete or categorical variables for data upload as highlighted in chapter 3. With the convergence of HDSS and other research studies data, the deployment of the application will now rely on the infrastructure.

Table 5: Research studies conducted in INDEPTH member HDSSs in Africa

List of research studies conducted in different HDD sites in Africa	
Research Study	HDSS Site(Country)
HIV/AIDS	Kisumu(Kenya), Dodowa(Ghana),Nouna(Burkina Faso)
Malaria	Kilifi(Kenya), Dodowa(Ghana), Kisumu(Kenya),Farafenni(Ghambia)
Verbal Autopsy	Kisumu(Kenya), Dodowa(Ghana), Argincourt(SA), Kilifi(Kenya)
Tuberculosis	Kisumu(Kenya), Dodowa(Ghana), AHRI(SA), Kilifi(Kenya)
Malnutrition	Kisumu(Kenya),Kilifi(Kenya),Nauna(Bukina Faso)
Pneumonia	Kisumu(Kenya),Kilifi(Kenya),Harar(Ethiopia)

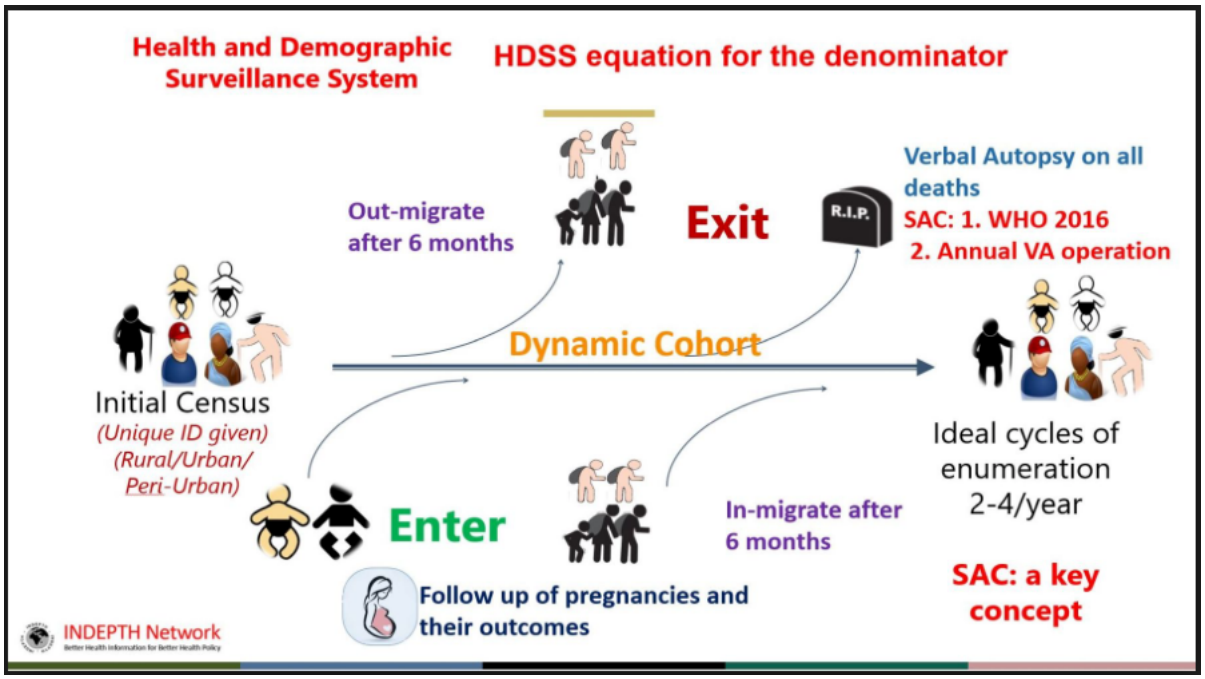


Figure 13: Health and Demographic Surveillance System linkage to Cohort studies

5.3 INFRASTRUCTURE CONSIDERATION

The main application is developed using rmarkdown and embedded on the shiny app and can be deployed to any major Operating System (OS) (Windows, Mac or Ubuntu) either in the server or client environment. The code is hosted at Github through which it provides source code control service and can be reused to promote knowledge sharing with accompanying documentation on how to host the application.

There are several ways in which the developed platform can be deployed, the fastest route would be to host the platform on Shiny cloud server. The cloud deployment is

scalable and easy to use with no hardware requirement or purchase contract required. However, for private shiny-server the following are the requirements;

- Installation of Linux distribution operating system e.g Ubuntu 14.04 and higher or Cent OS
- Latest version of R installed (Version R 3.5.2)
- Installation of RStudio server.

System requirement include;

- Minimum of 4 GB RAM; The amount of memory consumed by the application depends on the data loaded, thus with the ever-growing HDSS data it would be ideal to invest in scalable memory architecture.
- Since R is a single-threaded application it is recommended to have at least two running cores running at 2.0 GHz or greater.
- At least 20 GB solid-state drive (SSD) for reliability.

After downloading the application from Github and changing the working directory (using the setwd R function) for linkage to the dependency applications and data inputs(shapefile and vaccine data) as documented on the code (data files must be on the same directory as R Markdown document and Shiny application). First deploment after launching the application may take few minutes, as RStudio installs referenced packages to the new environment while creating other dependencies like folders for deployed scripts, functions and images. The deployment of the application will allow working from any internet enabled location after its configuration. For further references on administrators guide for installation of shiny server, see the Shiny server documentation.

5.4 SUMMARY

After the deployment of the platform to other HDSS centers, users can exploit the functions of the application and also engage in collaborative improvement with the implementation of important avenues for future work. Collaborations is hard, nevertheless with the use of Git for version control and Github which is ideal for decentralized repositories, task management, bug tracking, and future request will be easy to handle instead of exclusive ownership of source code.

6

CONCLUSION AND FUTURE DIRECTIONS

6.1 CONCLUSION

Dynamic visualization platforms remain a promising alternative for communication of complex datasets and displays, allowing health researchers to efficiently use and reason through the research data. To drive interest and support visualization applications, studies have highlighted the importance of on-going training for users to improve perceived transparency on the tool and which may increase the adoption rates [59]. Through the use of KHDSS data, we have shown that visual analytics tools can address the concerns of the populace.

Findings from the platform provided important insights and highlights the necessity for vaccinations of infants at the right time. Many Eyes [60], a web-based visual intelligent application tool identified polio vaccination coverage for children 19-35 months in the United States of America (USA), by race, poverty level and more importantly by geographic location. The application is been used to identify states with lower polio vaccine coverage and has helped to monitor polio in the USA with the last case occurring in 1979 [60]. These highlights the urgent need for policymakers to enhance monitoring, planning and promote recommended vaccination schedules for infants [61].

Infants born to mothers with a gestation period greater than 37.5 weeks, had a 9% reduced odds of dying. In sub-Saharan Africa, low birth weight due to preterm births is a predictor of infant mortality [62]. Tableau, health care analytics has been useful in identifying mortality risk factors, trends that exist and who is a risk [62].Our findings

suggest intervention strategies targeting preterm births to improve survival by scaling the use of visual analytics applications.

Surprisingly, infants born in hospital had a higher odds of dying compared to infants born at the homestead. This result might be confounded by several factors including: Fertility behavior, nutritional status, breastfeeding, infant feeding, health-seeking behaviors, environmental factors, health conditions, social-economic status, HIV status, and number of hospital admissions [63].

Visual analytics is an emerging field in research, and we demonstrated its strength with focus on preprocessing of data, geospatial analytics, statistical analytics and scientific analytics by providing smart hypothesis through knowledge discovery that will help to inform decision.

6.2 FUTURE DIRECTION

Future strategies of the visualization platform will focus on three main areas: First, the platform could benefit users by inclusion of a view presenting any questionable data in-order to draw meaningful and accurate conclusion. This can be achieved by including data quality function to avert any concerns. Skewed distribution, outliers, missing data, duplicate occurrences and in-consistency between relevant fields should be captured and allow filtering of such cases.

Another important feature is to make the application compatible with existing data formats and a feature to allow merging of different datasets seamlessly. To avoid redundancy of efforts in demographic surveillance sites, HDSS team collaborations is essential for the development of the application on a broader context of existing data.

The third future direction is integrating a model for predictive analytics. Depending on the selected target attribute, the platform should be able to predict future events like disease outbreaks, predict trends from time-series data on the surveillance system or even

predict individuals who as more likely to be infected. Statistical details for the model that provide details for the statistical test used with its effect size and statistical significate level will be provided for users on the dashboard.

A

PLAGIARISM DECLARATION

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG | FACULTY OF HEALTH SCIENCES

PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I __David Kivuli Amadi_____(Student number: _____969197_____) am a student registered for the degree of __MSc In Public Health Informatics_____in the academic year __2_____.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: _____Date: __28/03/2019_____

26/04/2015
1

Figure 14: Plagiarism Declaration

B

ETHICS CLEARANCE CERTIFICATE

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Mr David Amadi

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M171178

NAME: (Principal Investigator)	Mr David Amadi
DEPARTMENT:	Epidemiology and Biostatistics Kemri Welcome TRust Kilifi-Kenya
PROJECT TITLE:	An Integrated Knowledge Ecosystem for Infant Mortality in Kilifi Health and Demographic Surveillance (KHDSS)
DATE CONSIDERED:	24/11/2017
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Dr Gideon Nimako
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	18/09/2018

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary, Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in **November** and will therefore be due in the month of **November** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Figure 15: Ethics Clearance Certificate

C

DEPLOYMENT OF THE PLATFORM

Getting Started

These instructions will get you a copy of the project up and running on your local machine/server for development and testing purposes.

Prerequisites

- Fork the GitHub repository (https://github.com/daakiv/infant_mortality_app) i.e. allows you to copy the code without affecting the original.
- Clone the forked repository to your local system or server.
- Add Git remote to the original repository and place a branch on which to place changes.

Installing

- Change the working directory (using the setwd R function) for linkage to the dependency applications and data inputs (shapefile and vaccine data) as documented on the code (data files must be on the same directory as R Markdown document and Shiny application)

Deployment

- Host the platform on Shiny cloud server or deploy to private R studio server (see instructions for deployment at this link)
- Hosting on cloud Shinyapps.io is easy, secure and scalable with no hardware and software requirement. Free and paid options available.
- The deployment of the application will allow working from any internet enabled location after its configuration.
- Upload data for analysis with graphical display by point and click
- Use this link to view or use the application deployed for Kemri-Wellcome Trust Kilifi HDSS.

Built With

- shiny - The web framework used
- rmarkdown - Dependency Management

Versioning

We used Git and GitHub and for versioning.

Figure 16: Deployment of the Platform

BIBLIOGRAPHY

[1] Haidong Wang, Zulfiqar A Bhutta, Matthew M Coates, Megan Coggeshall, Lalit Dandona, Khassoum Diallo, Elisabeth Barboza Franca, Maya Fraser, Nancy Fullman, Peter W Gething, et al. "Global, regional, national, and selected subnational levels of stillbirths, neonatal, infant, and under-5 mortality, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015." In: *The Lancet* 388.10053 (2016), p. 1725.

[2] Li Liu, Shefali Oza, Dan Hogan, Yue Chu, Jamie Perin, Jun Zhu, Joy E Lawn, Simon Cousens, Colin Mathers, and Robert E Black. "Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals." In: *The Lancet* 388.10063 (2016), pp. 3027–3035.

[3] World Health Organization. *World health statistics 2016: monitoring health for the SDGs sustainable development goals*. World Health Organization, 2016.

[4] Francis S Collins. "What we do and don't know about'race','ethnicity', genetics and health at the dawn of the genome era." In: *Nature genetics* 36 (2004).

[5] J Ties Boerma and A Elisabeth Sommerfelt. "Demographic and health surveys (DHS: contributions and limitations." In: (1993).

[6] J Anthony G Scott, Evasius Bauni, Jennifer C Moisi, John Ojal, Hellen Gatakaa, Christopher Nyundo, Catherine S Molyneux, Francis Kombe, Benjamin Tsofa, Kevin Marsh, et al. "Profile: the Kilifi health and demographic surveillance system (KHDSS)." In: *International journal of epidemiology* 41.3 (2012), pp. 650–657.

[7] Philip Bejon, John Lusingu, Ally Olotu, Amanda Leach, Marc Lievens, Johan Vekemans, Salum Mshamu, Trudie Lang, Jayne Gould, Marie-Claude Dubois, et al. "Efficacy of RTS, S/AS01E vaccine against malaria in children 5 to 17 months of age." In: *New England Journal of Medicine* 359.24 (2008), pp. 2521–2532.

[8] Harvard Dataverse. *Harvard Dataverse Repository*. https://dataverse.harvard.edu/. Online; accessed 15-October-2018. 2019.

[9] Carren Ginsburg, Philippe Bocquier, Donatien Béguy, Sulaimon Afolabi, Orvalho Augusto, Karim Derra, Kobus Herbst, Bruno Lankoande, Frank Odhiambo, Mark Otiende, et al. "Healthy or unhealthy migrants? Identifying internal migration effects on mortality in Africa using health and demographic surveillance systems of the INDEPTH network." In: *Social Science & Medicine* 164 (2016), pp. 59–73.

[10] Carolyne Ndila, Evasius Bauni, George Mochamah, Vysaul Nyirongo, Alex Makazi, Patrick Kosgei, Benjamin Tsofa, Gideon Nyutu, Anthony Etyang, Peter Byass, et al. "Causes of death among persons of all ages within the Kilifi Health and Demographic Surveillance System, Kenya, determined from verbal autopsies interpreted using the InterVA-4 model." In: *Global health action* 7.1 (2014), p. 25593.

[11] Odette Wegwarth, Lisa M Schwartz, Steven Woloshin, Wolfgang Gaissmaier, and Gerd Gigerenzer. "Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States." In: *Annals of internal medicine* 156.5 (2012), pp. 340–349.

[12] Arjun K Manrai, Gaurav Bhatia, Judith Strymish, Isaac S Kohane, and Sachin H Jain. "Medicine's uncomfortable relationship with math: calculating positive predictive value." In: *JAMA internal medicine* 174.6 (2014), pp. 991–993.

[13] DeGaspari J. *Managing the Data Explosion*. https://www.hcinnovationgroup.com/interoperability-hie/cloud-computing-storage/article/13022085/managing-the-data-explosion. Online; accessed 15-October-2018. 2019.

[14] WHO. *Infant mortality*. Ed. by World Health Organizationl. URL: http://www.who.int/gho/child_health/mortality/neonatal_infant_text/en/.

[15] Jeffrey D Sachs. "From millennium development goals to sustainable development goals." In: *The Lancet* 379.9832 (2012), pp. 2206–2211.

[16] Kent Buse and Sarah Hawkes. "Health in the sustainable development goals: ready for a paradigm shift?" In: *Globalization and health* 11.1 (2015), p. 13.

[17] Aleksandra Jakubowski, Sally C Stearns, Margaret E Kruk, Gustavo Angeles, and Harsha Thirumurthy. "The US President's Malaria Initiative and under-5 child mortality in sub-Saharan Africa: A difference-in-differences analysis." In: *PLoS medicine* 14.6 (2017), e1002319.

[18] Kenneth Hill. "Frameworks for studying the determinants of child survival." In: *Bulletin of the World Health Organization* 81 (2003), pp. 138–139.

[19] Allan M Parnell, National Research Council (US) Committee on Population, et al. "Mechanisms for the Association of Maternal Age, Parity, and Birth Spacing with Infant Health." In: (1989).

[20] Abul Kashem Majumder, Marian May, and Prakash Dev Pant. "Infant and child mortality determinants in Bangladesh: Are they changing?" In: *Journal of Biosocial Science* 29.4 (1997), pp. 385–399.

[21] Fanciele Dinis Ribeiro, Rosângela Aparecida Pimenta Ferrari, Flávia Lopes Sant'Anna, José Carlos Dalmas, and Edmarlon Girotto. "Extremes of maternal age and child mortality: analysis between 2000 and 2009." In: *Revista Paulista de Pediatria* 32.4 (2014), pp. 381–388.

[22] Michael A Koenig, James F Phillips, Oona M Campbell, and Stan D'Souza. "Birth intervals and childhood mortality in rural Bangladesh." In: *Demography* 27.2 (1990), pp. 251–265.

[23] Kenya Demographic. "Health Survey 2014: key indicators." In: *Kenya National Bureau of Statistics (KNBS) and ICF Macro* (2014).

[24] Wendy Graham, Susannah Woodd, Peter Byass, Veronique Filippi, Giorgia Gon, Sandra Virgo, Doris Chou, Sennen Hounton, Rafael Lozano, Robert Pattinson, et al. "Diversity and divergence: the dynamic burden of poor maternal health." In: *The Lancet* 388.10056 (2016), pp. 2164–2175.

[25] Sunday Adepoju Adedini. "Contextual determinants of infant and child mortality in Nigeria." PhD thesis. 2014.

[26] Edmund Wedam Kanmiki, Ayaga A Bawah, Isaiah Agorinya, Fabian S Achana, John Koku Awoonor-Williams, Abraham R Oduro, James F Phillips, and James Akazili. "Socio-economic and demographic determinants of under-five mortality in rural northern Ghana." In: *BMC international health and human rights* 14.1 (2014), p. 24.

[27] Hisham Elmahdi Mustafa and Clifford Odimegwu. "Socioeconomic determinants of infant mortality in Kenya: analysis of Kenya DHS 2003." In: ().

[28] Michael Marmot. "Social determinants of health inequalities." In: *The lancet* 365.9464 (2005), pp. 1099–1104.

[29] WHO GIVS. *Global Immunization Vision and Strategy: 2006-2015. Geneva: WHO, 2005*. Tech. rep. Document WHO/IVB/05.05.

[30] Sudhir Anand and Till Bärnighausen. "Health workers and vaccination coverage in developing countries: an econometric analysis." In: *The Lancet* 369.9569 (2007), pp. 1277–1285.

[31] WHO Kenya. *WHO and UNICEF estimates of immunization coverage: 2015 revision*. 2016.

[32] Marie C McCormick. "The contribution of low birth weight to infant mortality and childhood morbidity." In: *New England journal of medicine* 312.2 (1985), pp. 82–90.

[33] Ingrid Waldron. "Sex differences in human mortality: the role of genetic factors." In: *Social science & medicine* 17.6 (1983), pp. 321–333.

[34] W Henry Mosley and Lincoln C Chen. "An analytical framework for the study of child survival in developing countries." In: *Population and development review* 10 (1984), pp. 25–45.

[35] Godelieve Masuy-Stroobant. "The determinants of infant mortality: how far are conceptual frameworks really modelled?" In: *The explanatory power of models*. Springer, 2002, pp. 15–30.

[36] Laurence SLUTSKER, Robert D NEWMAN, Abdisalan M NOOR, Juliette J MUTHEU, Andrew J TOTEM, Simon I HAY, and Robert W SNOW. "Insecticide-treated net coverage in Africa: mapping progress in 2000-07. Commentary." In: *Lancet* 373.9657 (2009).

[37] Lauren N Carroll, Alan P Au, Landon Todd Detwiler, Tsung-chieh Fu, Ian S Painter, and Neil F Abernethy. "Visualization and analytics tools for infectious disease epidemiology: a systematic review." In: *Journal of biomedical informatics* 51 (2014), pp. 287–298.

[38] Kenneth KH Chui, Julia B Wenger, Steven A Cohen, and Elena N Naumova. "Visual analytics for epidemiologists: understanding the interactions between age, time, and disease with multi-panel graphs." In: *PloS one* 6.2 (2011), e14683.

[39] Tom Koch and Kenneth Denike. "Crediting his critics' concerns: Remaking John Snow's map of Broad Street cholera, 1854." In: *Social science & medicine* 69.8 (2009), pp. 1246–1251.

[40] Robert Eugene Hoyt, Dallas Snider, Carla Thompson, and Sarita Mantravadi. "IBM Watson analytics: automating visualization, descriptive, and predictive statistics." In: *JMIR public health and surveillance* 2.2 (2016).

[41] JW Buehler. "CDC's vision for public health surveillance in the 21st century. Introduction." In: *MMWR supplements* 61.3 (2012), p. 1.

[42] Peter Fox and James Hendler. "Changing the equation on scientific data visualization." In: *Science* 331.6018 (2011), pp. 705–708.

[43] Tableau Software. *Make your data make an impact*. URL: https://www.tableau.com.

[44] Ayushi Shukla and Saru Dhir. "Tools for Data Visualization in Business Intelligence: Case Study Using the Tool Qlikview." In: *Information Systems Design and Intelligent Applications*. Springer, 2016, pp. 319–326.

[45] Ross Maciejewski, Stephen Rudolph, Ryan Hafen, Ahmad Abusalah, Mohamed Yakout, Mourad Ouzzani, William S Cleveland, Shaun J Grannis, and David S Ebert. "A visual analytics approach to understanding spatiotemporal hotspots." In: *IEEE Transactions on Visualization and Computer Graphics* 16.2 (2010), pp. 205–220.

[46] Karsten Tabelow, Jonathan D Clayden, P Lafaye De Micheaux, Jörg Polzehl, Volker J Schmid, and Brandon Whitcher. "Image analysis and statistical inference in neuroimaging with R." In: *NeuroImage* 55.4 (2011), pp. 1686–1693.

[47] Marco D'Ambros, Michele Lanza, and Martin Pinzger. "The metabase: Generating object persistency using meta descriptions." In: *Proceedings of FAMOOSR 2007* (2007).

[48] Leishi Zhang, Andreas Stoffel, Michael Behrisch, Sebastian Mittelstadt, Tobias Schreck, René Pompl, Stefan Weber, Holger Last, and Daniel Keim. "Visual analytics for the big data era—A comparative review of state-of-the-art commercial systems." In: *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE. 2012, pp. 173–182.

[49] Roger D Peng. "Reproducible research and biostatistics." In: *Biostatistics* 10.3 (2009), pp. 405–408.

[50] Leslie Curry. "The Future of the Public's Health in the 21st Century." In: *Generations* 29.2 (2005), p. 82.

[51] Roger D Peng, Francesca Dominici, and Scott L Zeger. "Reproducible epidemiologic research." In: *American journal of epidemiology* 163.9 (2006), pp. 783–789.

[52] Osman Sankoh and Peter Byass. *The INDEPTH Network: filling vital gaps in global epidemiology*. 2012.

[53] Kobus Herbst, Sanjay Juvekar, Tathagata Bhattacharjee, Martin Bangha, Nidhi Patharia, Titus Tei, Brendan Gilbert, and Osman Sankoh. "The INDEPTH Data Repository: An International Resource for Longitudinal Population and Health Data From Health and Demographic Surveillance Systems." In: *Journal of Empirical Research on Human Research Ethics* 10.3 (2015), pp. 324–333.

[54] Martin Bangha, Alioune Diagne, Ayaga Bawah, and Osman Sankoh. "Monitoring the millennium development goals: the potential role of the INDEPTH Network." In: *Global health action* 3.1 (2010), p. 5517.

[55] Justus Benzler, Kobus Herbst, and Bruce MacLeod. *A data model for demographic surveillance systems*. 1998.

[56] Margaret Gyapong, Doris Sarpong, Elizabeth Awini, Alfred K Manyeh, Desmond Tei, Gabriel Odonkor, Irene A Agyepong, Precious Mattah, Peter Wontuo, Mary Attaa-Pomaa, et al. "Profile: the Dodowa HDSS." In: *International journal of epidemiology* 42.6 (2013), pp. 1686–1696.

[57] Frank O Odhiambo, Kayla F Laserson, Maquins Sewe, Mary J Hamel, Daniel R Feikin, Kubaje Adazu, Sheila Ogwang, David Obor, Nyaguara Amek, Nabie Bayoh, et al. "Profile: the KEMRI/CDC health and demographic surveillance system—Western Kenya." In: *International journal of epidemiology* 41.4 (2012), pp. 977–987.

[58] Kathleen Kahn, Mark A Collinson, F Xavier Gómez-Olivé, Obed Mokoena, Rhian Twine, Paul Mee, Sulaimon A Afolabi, Benjamin D Clark, Chodziwadziwa W Kabudula, Audrey Khosa, et al. "Profile: Agincourt health and socio-demographic surveillance system." In: *International journal of epidemiology* 41.4 (2012), pp. 988–1001.

[59] Anthony C Robinson, Alan M MacEachren, and Robert E Roth. "Designing a web-based learning portal for geographic visualization and analysis in public health." In: *Health informatics journal* 17.3 (2011), pp. 191–208.

[60] Richard Higgins, Guisu Li, and Richard Ytzen Tahira Ahmad. "Visualizing Polio Vaccination Trends in the United States: Status of the Nation." In: ().

[61] Sidu Biai, Amabelia Rodrigues, Jens Nielsen, Morten Sodemann, and Peter Aaby. "Vaccination status and sequence of vaccinations as risk factors for hospitalisation among outpatients in a high mortality country." In: *Vaccine* 29.20 (2011), pp. 3662–3669.

[62] Joanne Katz, Anne CC Lee, Naoko Kozuki, Joy E Lawn, Simon Cousens, Hannah Blencowe, Majid Ezzati, Zulfiqar A Bhutta, Tanya Marchant, Barbara A Willey, et al. "Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis." In: *The Lancet* 382.9890 (2013), pp. 417–425.

[63] Shea O Rutstein. "Factors associated with trends in infant and child mortality in developing countries during the 1990s." In: *Bulletin of the World Health Organization* 78 (2000), pp. 1256–1270.