Kibuon Data Analysis: Determining an Optimal Well Location and Water Usage Trends for a Community in the Southwest Region of Kenya

I. ABSTRACT – Firsthand data collected by a team of 8 in the southeast region of Kenya is used to create visualizations from which was usage and need trends can be inferred. The data is also used to determine an optimal location for installation of a well for water provision.

II. INTRODUCTION

Kibuon is a community in the southwest region of Kenya. The community consist of about 1600 people. They face challenges pertaining to water access: due to lack of nearby clean water sources, the community members have to travel long distances to collect water from springs. Furthermore, these springs replenish slowly leading to long wait times and the water is unclean leading to illnesses when consumed. The community members therefore partnered with one of the college chapters of Engineers Without Borders USA (an international humanitarian organization) to work on a long-term sustainable water provision project. The data used in this report is from an assessment trip conducted in the Summer of 2019 by a team of 6 engineering students, 1 profession civil engineering mentor and 1 engineering professor. The objective is to determine an optimal locations and number of wells to install in partnership with the community.

This report follows the following structure: Section III provides the background context on the geography of Kibuion. It is followed by a discussion of the sources of data and the tools used to parse, process and analyze the data in Section IV. Section V presents some graphs generated from some of the data and discusses the implications.

III. CONTEXT

The Kibuon community consist of four regions A(North West), B(North East), C(South East), D(South West) as shown below in Figure 1. The region has a total of about 200 houses.

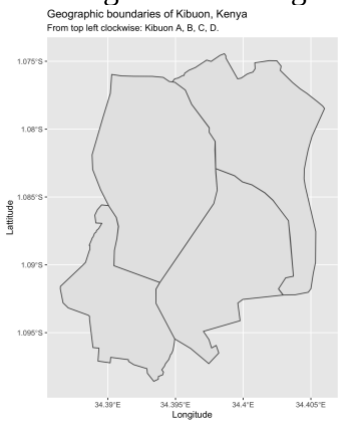


Figure 1: The four regions of Kibuon

While conducting the assessment trip, the community members were asked to give suggestions for potential locations for installing wells to which they responded with 8 different locations namely: Kibuon Maranatha, Kater, Munyu Maranatha, Kawegi, Ombolwanda, Shirikisho Church, Kadinda and one other unassessed location referred to herein as Unassesed. Part of that analysis done on this location data is presented in this report in section V.

IV. DATA SOURCES

This project uses three main data sources: the first is s KMZ geodata file and the second is a survey database accumulated from surveys conducted in the community during the assessment trip.

A. KMZ DATA

The KMZ file contains geographic boundary information such as the one used to make Figure 1 above. The KMZ file also consists of coordinates for all house in the community and the number of people in each of those houses. The demographic data allows for extrapolations to be done over the entire community for factors that can be modelled as functions of population.

This data was parsed in python to extract the coordinates and the number of people per house. It was then processed to make a regressions and plots in R.

B. SURVEY DATA

The survey team collected data from 40 households in the community. The selection of houses to survey was done on a random basis. In the survey, data corresponding to the following variables was collected: number of people living in the house, sources of water in the dry season, sources of water in the wet season, time taken to travel to water sources in the respective seasons, methods for travelling to, the water sources, challenges faced with regards to water collection, amount of water used per day in liters per household for general purposes (cleaning, washing, bathing, etc), amount of water used for drinking in liters per day per household, gender and age of the people who collect water for the household, methods of cleaning water for members of the household, symptoms of illnesses in the community, and cost for buying water from neighboring communities.

The data was available in csv format therefore it was directly imported and processed using R. The plots below show the distribution of surveyed houses(left) and the distribution of all the houses in the community(right)

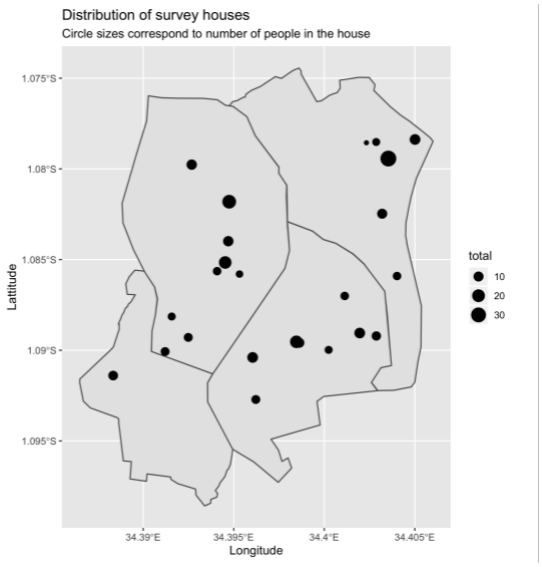


Figure 1: Surveyed houses

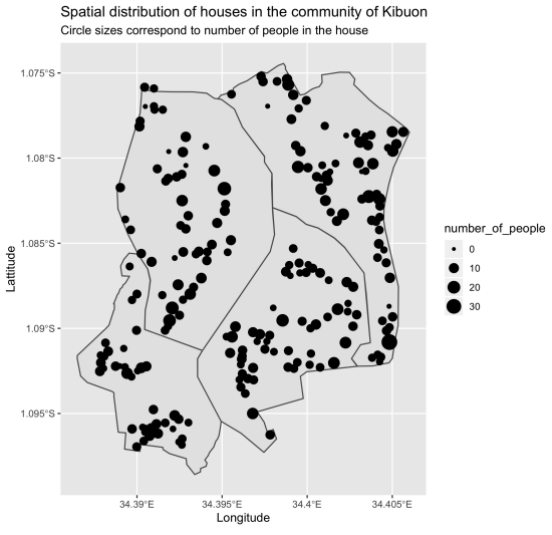


Figure 2: All the housed in the community

V. DATA ANALYSIS

A. Optimal Well Location based on distance distribution

The coordinates of all the house from the KMZ is used to determine the distance from each potential well location to all the houses. The results shown below in Figure 3 suggest that the optimal location for installing a well based on proximity to most of the houses is Kater then either Kibuon Maranatha or the Unassesed Location.

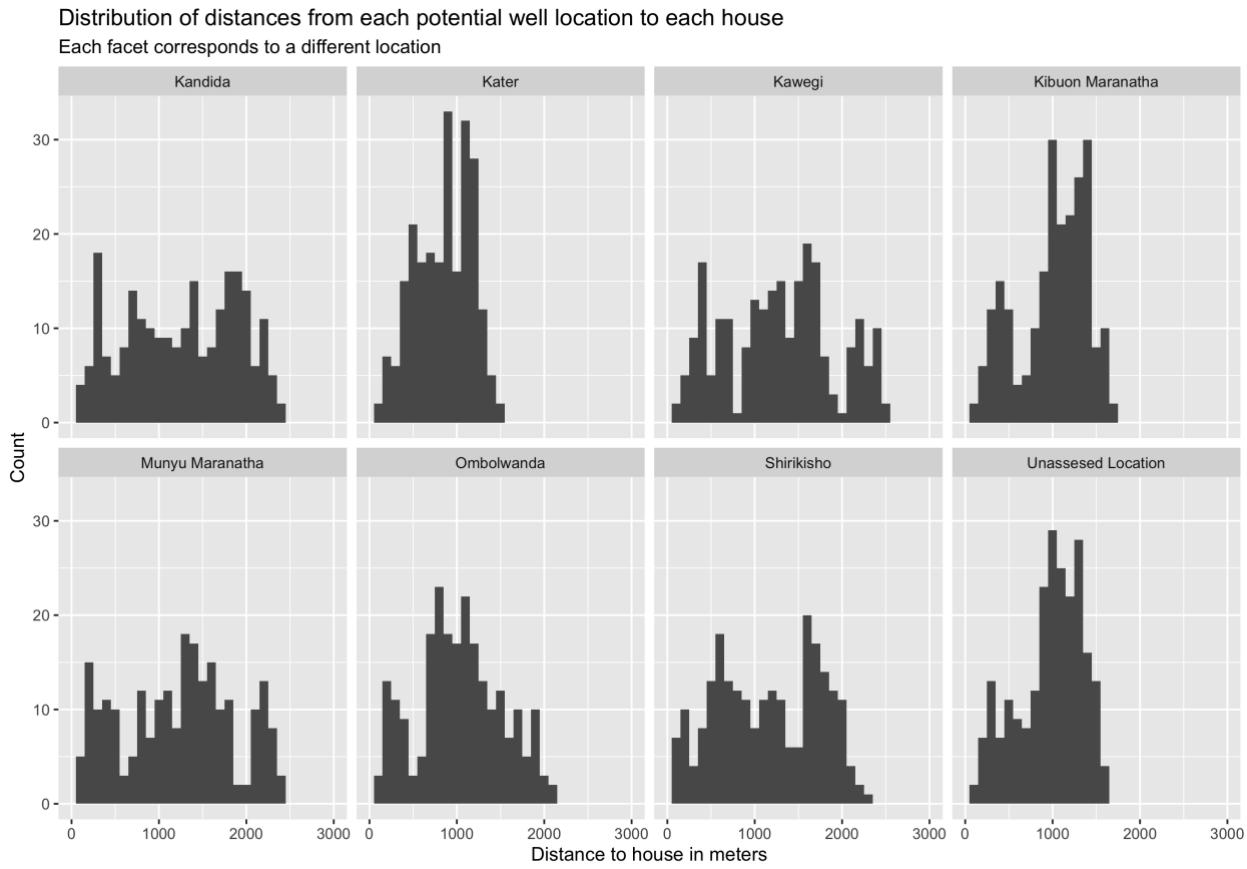
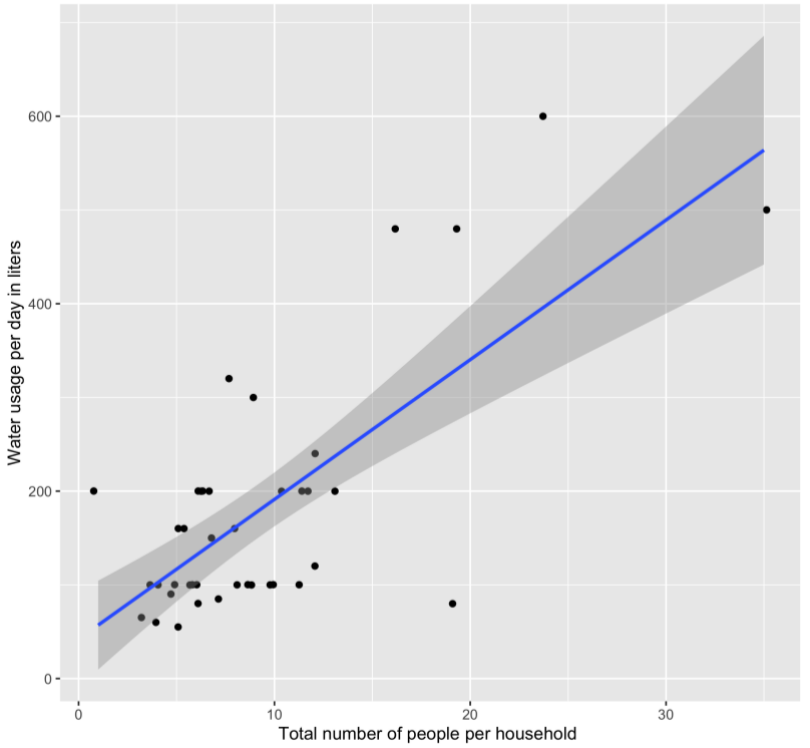


Figure 3: Distribution of distances from each well to all the houses in the community. The optimal locations based on proximity to houses is Kater and Kibuon Maranatha.

B. Estimation of water usage per household per day in the entire community

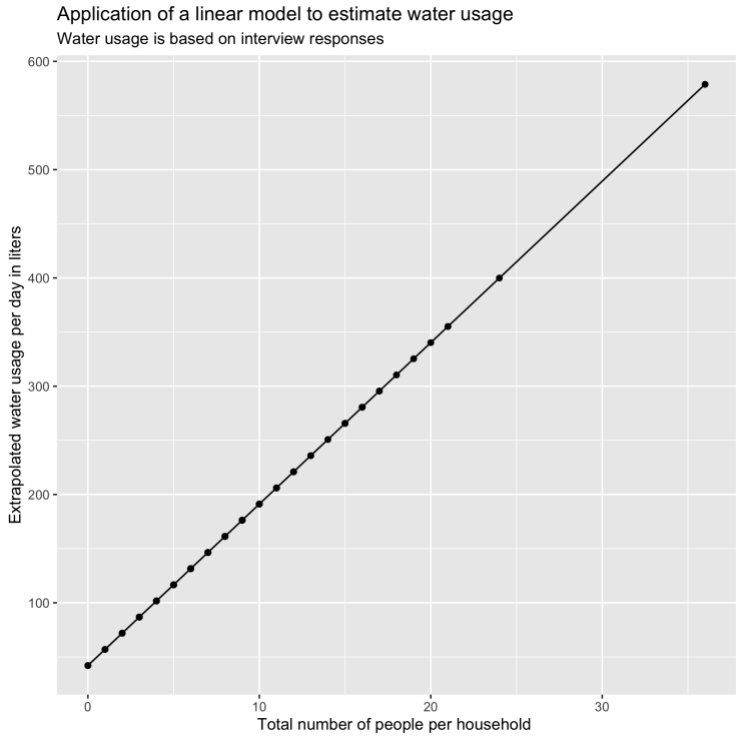
A linear regression is done on the survey data to model the amount of water used per household per day as a function of the total number of people in the house. A graph of this regression can be found in Appendix I. This model is then applied to the population data from the KMZ to extrapolate and estimate the water needs in the entire community. A graph of this extrapolation is included in Appendix II.

Appendix I



Linear model of water usage as a function of number of people

Appendix II



Extrapolation of the linear model over the population data of the whole community