

Pump it Up: Data Mining the Water Table in Tanzania

Stuart King

June 2, 2016

Agenda

- Introduction
 - About the data
 - Research question and key findings
- Overview of data handling
- Analysis
 - Location
 - Installers
 - Management structure
 - Payment schemes
- Summary of findings
- Recommendations and next steps

Introduction: About the Data

- Two datasets taken from DrivenData.org as part of their open data mining competition
- Data encompasses descriptive variables for 59,400 water pumps (“points”) throughout Tanzania
- Data compiled by Taarifa and the Tanzanian Ministry of Water
 - The Taarifa platform is an open source web API created during the 2012 World Bank-sponsored Sanitation Hackathon used to crowd source and triage infrastructure-related social issues



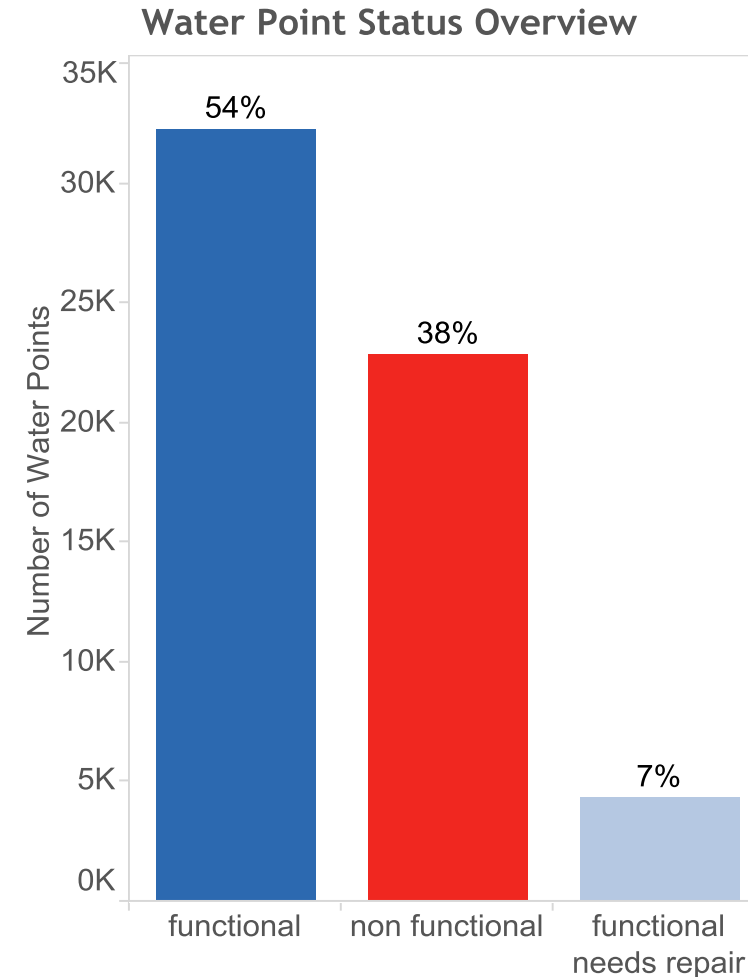
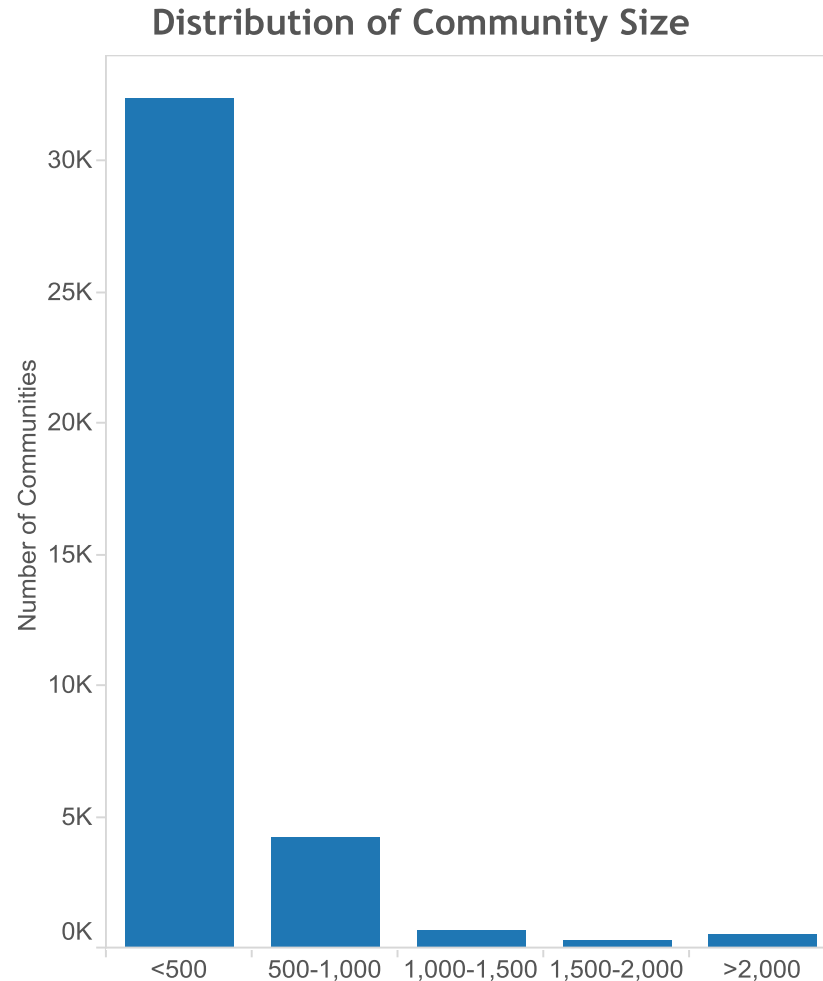
Introduction: Research Question and Findings

- Business scenario:
 - The client is preparing for an anticipated call for proposals from the World Bank to improve access to clean water in Tanzania. To inform its project proposal the client would like information about the functional status of water points throughout Tanzania.
- Research question(s):
 - What factors contribute to the functional status of water points in Tanzania, and where is the greatest need for water point rehabilitation?
- Key findings:
 1. Regions in the southeast have a greater proportion of non-functional water points
 2. DWE is the chief installer and VWCs are your managers
 3. Payment schemes work!

Overview of Data Handling

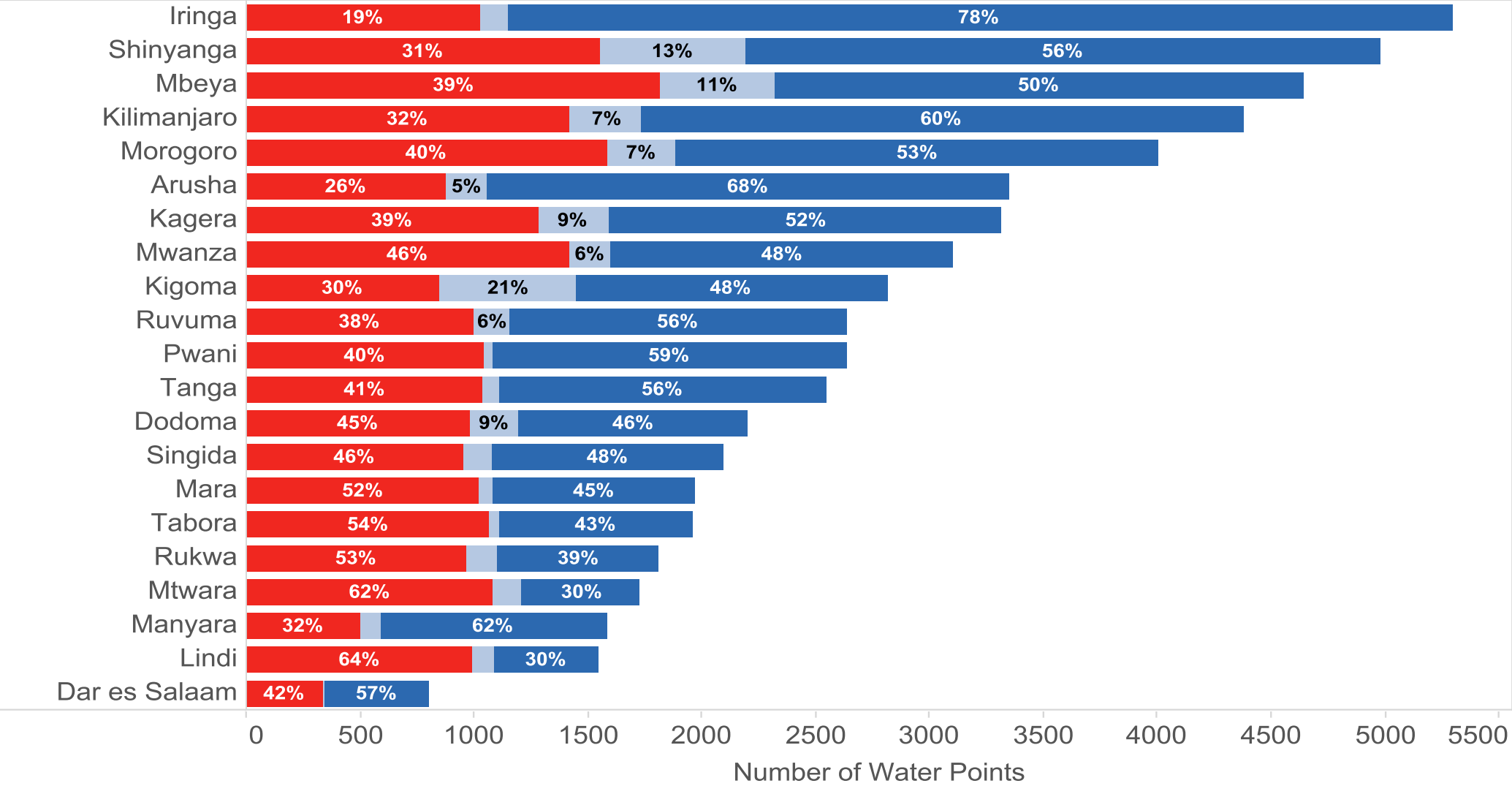
- Cleaned *installer*, *management*, and *region* variables to normalize values
- For analysis in Excel, combined the two datasets, matching water point id codes to assign a functional condition
- Created new variables to segment by full functionality and payment scheme
- Calculated the number of years since each water point was installed, imputing the average water point age for null values
- Used Tableau to create new fields to group data for distribution analysis
- Null values generally ignored, unless otherwise stated

Most water points are in small communities and almost half are either dysfunctional or need repair



Iringa leads the way!

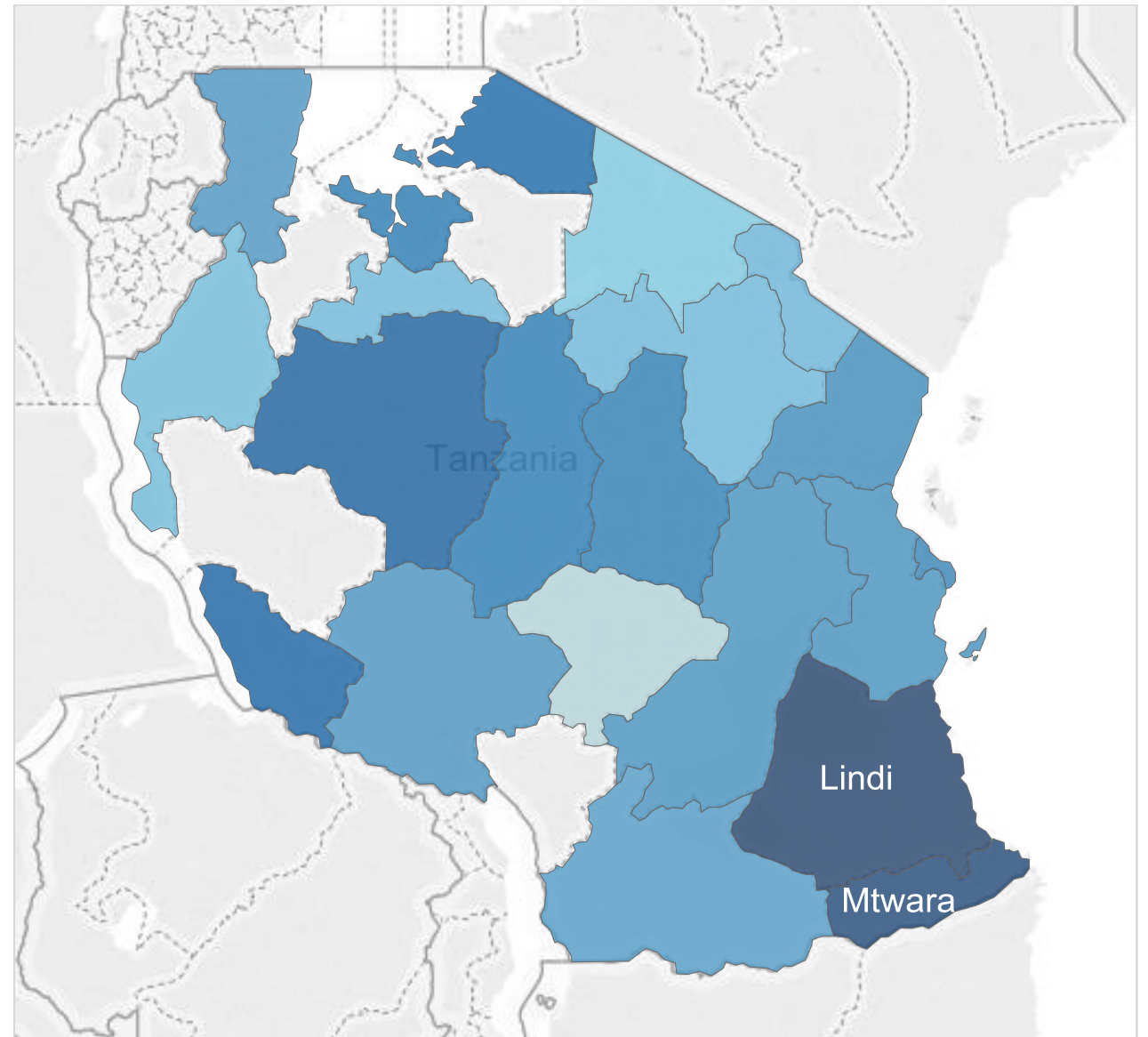
Water Points per Region, Total and Status %



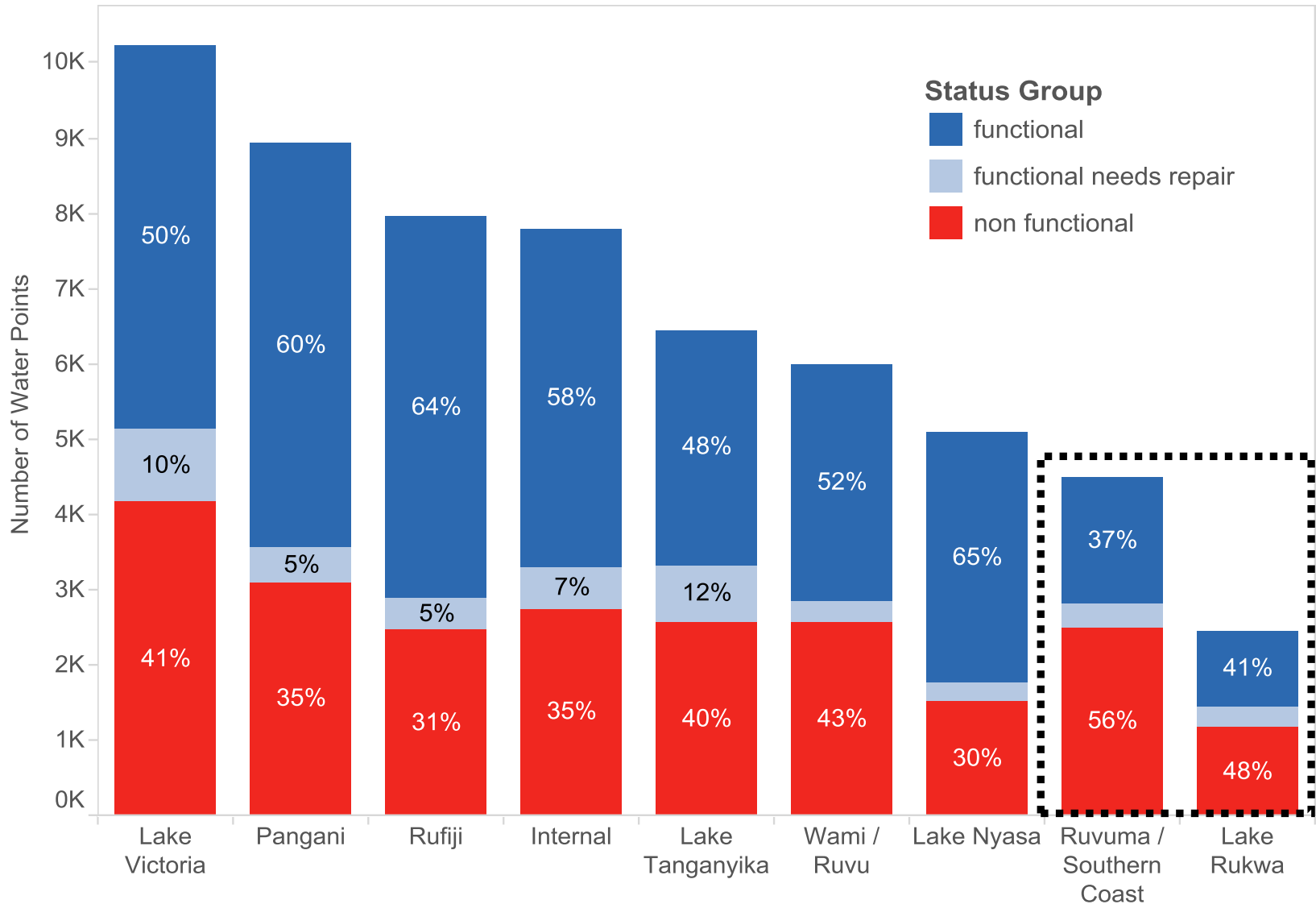
Regions in the **Southeast**
have a higher proportion of
non-functional pumps

Proportion of Functionality by Region

	functional	functional needs repair	non functional
Lindi	30%	6%	64%
Mtwara	30%	7%	62%
Tabora	43%	2%	54%
Rukwa	39%	7%	53%
Mara	45%	3%	52%



Functional Status of Water Points by Water Basin



Ruvuma/Southern Coast and Lake Rukwa river basins have a higher proportion of non-functional water points than fully functional water points

The Department of
Water Engineer
 (DWE) dominates the
 market; **80%** of
 Finwater-installed
 pumps are
 dysfunctional

Installer	# of water points	% of total	% of installed water points		
			functional	functional needs repair	non functional
DWE	17,844	30.0%	54%	9%	37%
Central Government	3,815	6.4%	33%	8%	59%
Danida	1,787	3.0%	60%	5%	35%
Community	1,676	2.8%	68%	5%	27%
Private	1,607	2.7%	77%	3%	20%
Hesawa	1,401	2.4%	56%	4%	40%
RWE	1,307	2.2%	26%	11%	63%
KKKT	1,207	2.0%	49%	6%	44%
District Council	1,172	2.0%	42%	7%	51%
Finwater	802	1.4%	16%	4%	80%

Time Since Water Point Construction - Top 10 Installers by Amount

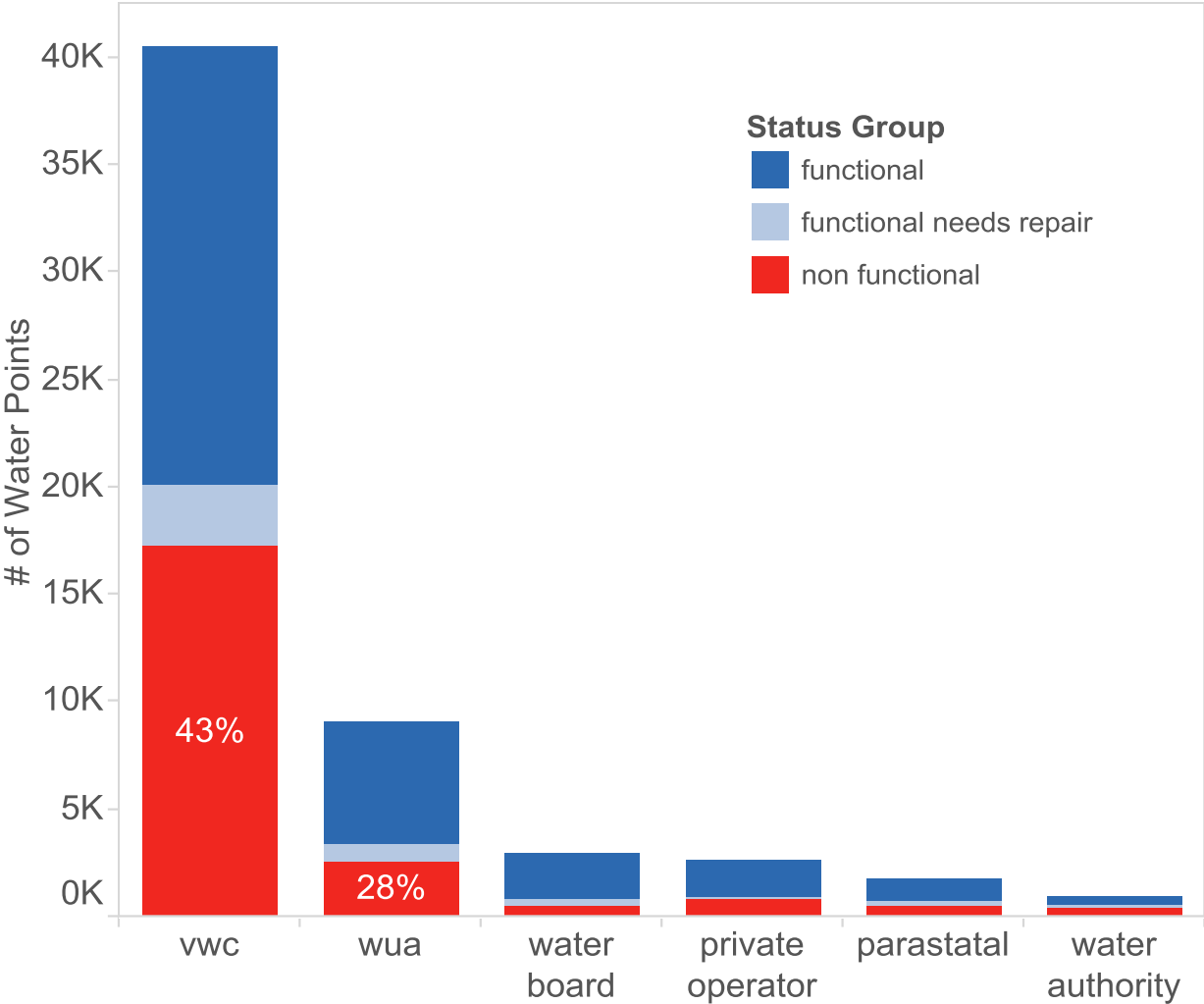
	Central Govnt	Community	Danida	District Council	DWE	Finwater	Hesawa	KKKT	Private	RWE
<10 yrs	530	571	84	447	4,010	20	6	134	951	14
10-20 yrs	150	256	406	126	2,943	17	357	83	201	129
20-30 yrs	108	233	905	189	1,652	250	96	4	39	288
30-40 yrs	341	95	113	214	1,733	457	13	25	21	561
>40 yrs	312	37		33	726	15	1	2	10	181

43% of water points managed by village water committees are non-functional

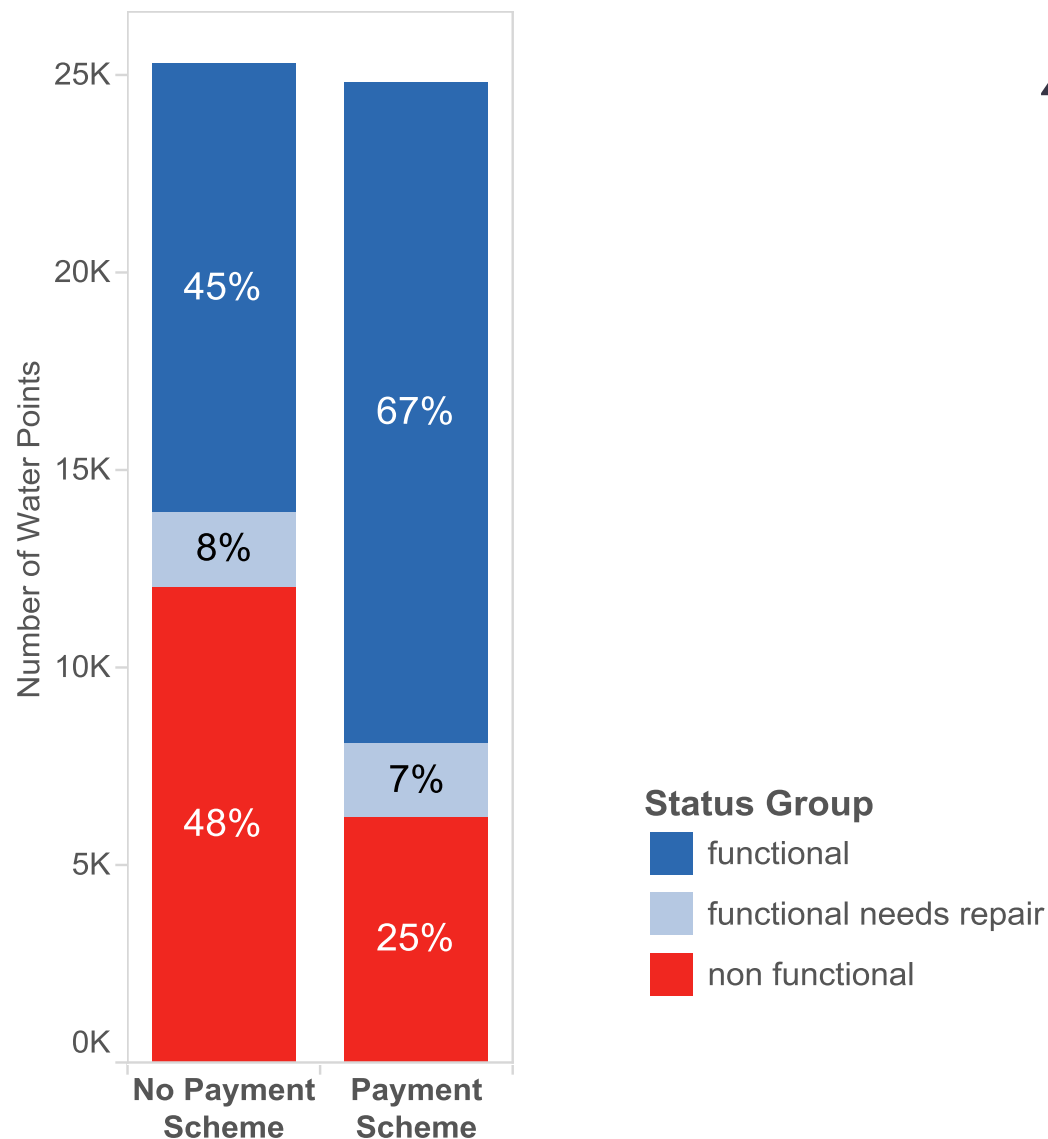
Functional Status Breakdown of Water Points per Management Structure

	functional	functional needs repair	non functional
water board	74%	9%	17%
private operator	66%	2%	32%
wua	63%	9%	28%
trust	59%	8%	33%
parastatal	58%	12%	30%
vwc	50%	7%	43%
water authority	49%	6%	45%
school	23%	1%	76%

Functional Status of Water Points by Management Structure



Functional Status of Water Points with and without Payment Schemes



48% of water points without a payment scheme are non-functional, compared to only 25% of water points with a payment scheme being non-functional

Summary of Key Findings

- 1) The **Lindi** and **Mtwara** regions along the southeast coast of the country have the greatest proportion of non-functional pumps
- 2) The **DWE** is by far the most active installer in Tanzania; Watch out for **Finwater!**
- 3) **VWCs** manage the vast majority of water points and **43%** of managed water points are non-functional
- 4) Water points with a payment scheme are **more likely** to be functional than pumps without a payment scheme

Recommendations & Next Steps

- Dependent variable (functional status) is categorical – perform logistical regression using dataset variables to enrich analysis
- Further research into payment schemes to determine what scheme works best
- Improving the skills and capacities of the DWE will increase the sustainability and quantity of fully functional water points
- VWC should be a target stakeholder group since they manage the majority of pumps
 - Understanding the social dynamics and how best to work with these groups will improve the technical proposal
- Deeper analysis into age of water points could improve predictability of functional status and likelihood of needed repairs