

# Uchaguzi Analysis and Research Overview

## What is Uchaguzi Analysis and Research team goal?

Making sense of the "dots" and amplifying the key data points is the root of the Analysis team. They will work on the tips and planning built by the Uchaguzi Analytics, Research and Analysis Community Working Group.

It is very hard to estimate the number of messages we will receive. The Communication Commission of Kenya latest figures show mobile telephony penetration at 75.4 per cent in an industry with a total of 29.7 million subscribers. What this means is that nearly everyone between 15 - 65 years old (the most likely to own a mobile phone) are connected. Together with our partners Uchaguzi will be rolling out an extensive advertising campaign to let the public know about the deployment and how to send information. On election day we may process 50,000 different pieces of information. In the days leading up to the general election perhaps 5000 - 10,000. However, these are guess.

Please see the Uchaguzi case study for some analytics on this. <http://www.slideshare.net/Ushahidi/kenya-ushahidi-evaluation-uchaguzi>  
<http://www.hivos.org/activity/ict-election-watch-uchaguzi-deployment-kenya-20122013>

IEBC (electoral commission) has released more details. This has been complex to get. Mikel has provided us with Kenyan shapefiles. We are also working with the UMATI project (which will handle the category - "hate speech") <http://www.ihub.co.ke/blog/2012/12/umati-online-media-monitoring-project-releases-november-2012-initial-results/>

## Uchaguzi Research and Analysis Team Process Manual

### 1.1 Scheduling:

- Log into the shared calendar to indicate the times you will be able to participate. <https://docs.google.com/spreadsheet/ccc?key=0Ajto4YrsWC3bdFlwZfVwWmIHTjh1QWcxYjRmRUcwOEE&usp=sharing>
- Log into the team Skype and remain online for communication purposes while you are working on the reports. This will be useful for regular update and constant communication with the rest of the team

### 1.2 Data

- We will be working mainly with verified data from uchaguzi.co.ke.
- Log in to the site with your unique log in details
- Open the the analysis log [https://docs.google.com/a/ihub.co.ke/spreadsheet/ccc?key=0AuQwHZnS\\_LGjdHZPU25hRONtUDYtNWh4em5SME1COGc&usp=sharing](https://docs.google.com/a/ihub.co.ke/spreadsheet/ccc?key=0AuQwHZnS_LGjdHZPU25hRONtUDYtNWh4em5SME1COGc&usp=sharing) to find the report number, date and time where previous volunteer left off (Click on link to request access)
- Download reports from the dashboard of uchaguzi.co.ke. Follow instructions [here](#) .
- Indicate to download reports from the time and date the previous volunteer left off up to the current time and date
- Once you have downloaded your dataset, log the report numbers you are working on and the time and dates of your report downloads in the analysis log document, to allow the next volunteer to pick up from where you left off.

### 1.3 Analysis

- Sift through the data and analyze for emerging trends, patterns, information gaps and critical and urgent issues to be addressed.
- There are a number of tools to be used in analysis of this data. The data will primarily be in csv formats and you can use excel as the basis for analysis. Other tools can also be used to further analyze this data and visualize it. However, kindly leave your outputs as Excel formats on in Word documents
- The analysis should be done as follows:
  1. According to locations and categories(if it falls under multiple categories, indicate all).
  2. Highlight the number of reports that you have worked on in order to draw your analyses.
  3. Leave your analyses, if possible, in tabular formats.
  4. Further, kindly write a short write-up of your analyses and key findings.
- These analyses should be uploaded, together with the raw data, into the shared analyses folders ( marked by dates [https://docs.google.com/a/ihub.co.ke/folder/d/0B-QwHZnS\\_LGjR1V4akVEUctOFU/edit?usp=sharing](https://docs.google.com/a/ihub.co.ke/folder/d/0B-QwHZnS_LGjR1V4akVEUctOFU/edit?usp=sharing))
  1. Upload you analyses either in excel format or a word doc that can easily be read from any machine.
  2. Label your files as follows: Name\_Time\_report number\_a\_b where 'a' is the first report in the raw data and 'b' is the last report in the raw data
- These uploaded analyses will enable the co-leads compile, cross compare and visualize your analyses.
- The number of reports should be indicated on your analysis doc and the log doc (as well as the range of reports you have worked on)

## 1.4 Outputs:

- Outputs from this team will include write-ups of research results in the form of regular blog posts and situation reports.
- These will be shared using the following schedule
  - March 3rd - 3:30pm report due (covering March 2nd 00:00 - 23:00 & March 3rd 00:00 - 15:00 EAT)
  - March 4th - 3:30 pm report due (covering March 3rd 15:00 - 23:00 & March 4th 00:00 - 15:00 EAT)
  - March 5th - 3:30 pm report due (covering March 4th 15:00 - 23:00 & March 5th 00:00 - 15:00 EAT)
  - March 6th - 3:30 pm report due (covering March 5th 15:00 - 23:00 & March 6th 00:00 - 15:00 EAT)

## Checklist/Summary for Research And Analysis Team

Indicate the times you will be working in the schedule: <https://docs.google.com/spreadsheet/ccc?key=0Ajto4YrsWC3bdFlwZfVaWmIHTjh1QWcxYjRmRUcwOEE&usp=sharing>

Log into Skype

Log into Uchaguzi.co.ke

Open the Analysis log to find out where previous volunteer left off: [https://docs.google.com/a/ihub.co.ke/spreadsheet/ccc?key=0AuQwHZnS\\_LGjdHZPU25hRONtUDYtNWh4em5SME1COGc&usp=sharing](https://docs.google.com/a/ihub.co.ke/spreadsheet/ccc?key=0AuQwHZnS_LGjdHZPU25hRONtUDYtNWh4em5SME1COGc&usp=sharing)

Download reports from where the previous volunteer left off to current date and time

Log the reports you are working on in the Analysis log

Analyze the reports

Upload your analyses and raw data in Excel/Word formats here: [https://docs.google.com/a/ihub.co.ke/folder/d/0B-QwHZnS\\_LGjR1V4akVEUnctOFU/edit?usp=sharing](https://docs.google.com/a/ihub.co.ke/folder/d/0B-QwHZnS_LGjR1V4akVEUnctOFU/edit?usp=sharing) Inform your lead that you've uploaded analysis and are checking out.

## Background Resources

### Tools

It is without any doubt that many of us, being professionals, scholars, and students, have had some experience in analyzing data and have used various tools for supporting that work. Please put down the list of tools that you believe will be useful in performing various statistical methods and other data-driven analytical techniques in the table below. The table will help us choose which tools will be adequate enough, given the expertise of this team, to carry out the right analyses.

Name of Tool	Description	Supported File Formats	Computer Platform Support	Link	Learning Curve	Notes
Microsoft Excel	Spreadsheet application for organizing, analyzing, and visualizing data in worksheets.	.xls, .xlsx, .csv, .txt	Mac and Windows	<a href="http://office.microsoft.com/en-us/excel/">http://office.microsoft.com/en-us/excel/</a>	low	Cost is around ~\$120 for Microsoft Office for Mac (Word, Excel, Powerpoint, Outlook).

Tableau	Desktop application that lets you easily import your data and offers easy mechanisms for visualizing it.	.xls, .txt, databases (MySQL, Oracle, PostgreSQL, etc.), Hadoop, etc.	Only Windows	<a href="http://www.tableausoftware.com/">http://www.tableausoftware.com/</a>	low	The Tableau personal edition can be used for free, which allows data sources like .xls and .txt. However, the next step up, professional edition, costs up to \$1999 per user. Finally, the Tableau Server edition allows anyone to publish her data on a server that can then be accessed on an Internet browser at a public URL. More info is found here: <a href="http://www.tableausoftware.com/products/desktop/specs">http://www.tableausoftware.com/products/desktop/specs</a>
Matlab	Desktop application and framework that puts together a programming language and various libraries for manipulating and visualizing data.	as far as I know (AFAIK), everything	Mac, Windows	<a href="http://www.mathworks.com/products/matlab/">http://www.mathworks.com/products/matlab/</a>	high	Though Matlab is a very powerful environment, it doesn't come without its costs (starts at ~\$500) and its steep learning curve (learning the programming language, acquaintance with toolsets and libraries, etc.).
Open Refine	A free, open source, power tool for working with messy data.	TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with OpenRefine extensions.	Mac, Windows and Linux	<a href="https://code.google.com/p/google-refine/wiki/Downloads">https://code.google.com/p/google-refine/wiki/Downloads</a>	low	Useful feature is the "text facet". Used to group related data from Ushahidi reports, eg categories.  Open Refine is easy to get started with and works on a web browser, though advanced features might take some time to learn.

Please share resources and best practices for analysis and research.