

# CONVERTING DATA TO FRIENDLIER FORMATS

Local voices. Global change.

# Contents

## Scraping data from the web

- Using browser plug-ins to scrape data tables
- Exercise
- Using Google docs to import data into charts
- Exercise

## Summary & Resources

# Contents

Scraping data from the web

- **Using browser plug-ins to scrape data tables**
- Exercise
- Using Google docs to import data into charts
- Exercise

Summary & Resources

# Scraping data using plug-ins

- Browser scrapers:
  - [Dafizilla Table2Clipboard](#) for Mozilla Firefox
  - [Scraper Extension](#) for Chrome
- Allow you to select tables on a website, including specific rows and columns
- Copy the data tables and either create a Google spreadsheet or paste into a blank Excel file

# Contents

## Scraping data from the web

- Using browser plug-ins to scrape data tables
- **Exercise**
- Using Google docs to import data into charts
- Exercise

## Summary & Resources

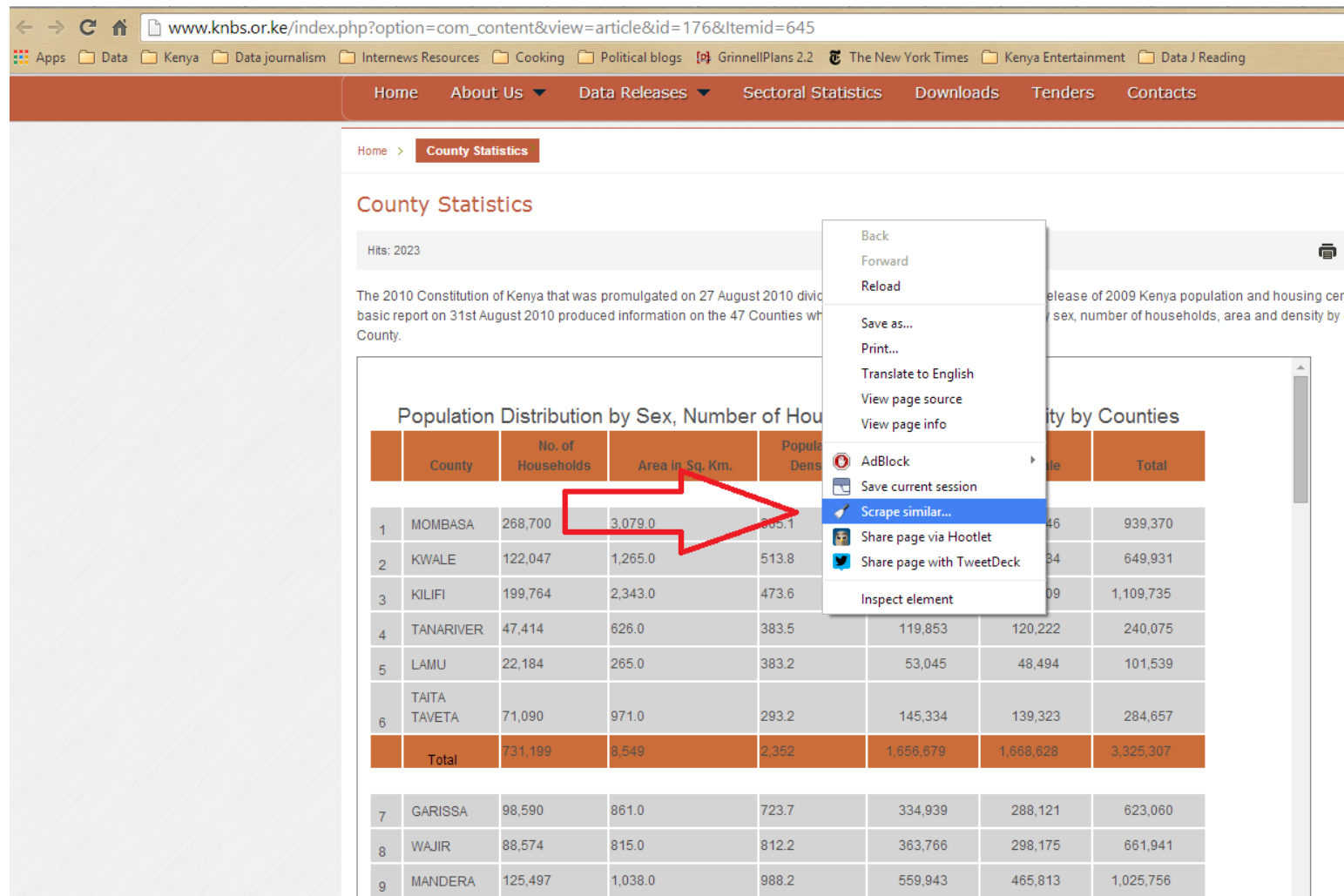
# Scraping data using browser extensions exercise

1. The appropriate browser extension should be installed:
  - [Dafizilla Table2Clipboard](#) for Mozilla Firefox
  - [Scraper Extension](#) for Chrome
2. Go to [KNBS County Statistics](#)
3. If you are using the Chrome, highlight part of the table, right click and select 'Scrape similar' then Export to Google docs.

If you are using Mozilla, right click on the table and select "Table2Clipboard" and copy whole table. Then open an Excel file and paste.



# Scraping data using browser extensions exercise



www.knbs.or.ke/index.php?option=com\_content&view=article&id=176&Itemid=645

Apps Data Kenya Data journalism Internews Resources Cooking Political blogs GrinnellPlans 2.2 The New York Times Kenya Entertainment Data J Reading

Home About Us Data Releases Sectoral Statistics Downloads Tenders Contacts

Home > County Statistics

### County Statistics

Hits: 2023

The 2010 Constitution of Kenya that was promulgated on 27 August 2010 divided the country into 47 counties. The basic report on 31st August 2010 produced information on the 47 Counties with the following information: County.

#### Population Distribution by Sex, Number of Households and Area in Sq. Km. by Counties

	County	No. of Households	Area in Sq. Km.	Population Density	Male	Female	Total
1	MOMBASA	268,700	3,079.0	85.1	128,466	140,234	939,370
2	KWALE	122,047	1,265.0	513.8	61,023	61,024	649,931
3	KILIFI	199,764	2,343.0	473.6	99,882	99,882	1,109,735
4	TANARIVER	47,414	626.0	383.5	119,853	120,222	240,075
5	LAMU	22,184	265.0	383.2	53,045	48,494	101,539
6	TAITA TAVETA	71,090	971.0	293.2	145,334	139,323	284,657
	<b>Total</b>	<b>731,199</b>	<b>8,549</b>	<b>2,352</b>	<b>1,656,679</b>	<b>1,668,628</b>	<b>3,325,307</b>
7	GARISSA	98,590	861.0	723.7	334,939	288,121	623,060
8	WAJIR	88,574	815.0	812.2	363,766	298,175	661,941
9	MANDERA	125,497	1,038.0	988.2	559,943	465,813	1,025,756

# Contents

## Scraping data from the web

- Using browser plug-ins to scrape data tables
- Exercise
- **Using Google docs to import data into charts**
- Exercise

## Summary & Resources



# Scraping data using Google spreadsheets and Import HTML

1. Open  
[http://en.wikipedia.org/wiki/China\\_at\\_the\\_Olympics](http://en.wikipedia.org/wiki/China_at_the_Olympics)
2. Open a Google Spreadsheet
3. Type in  
`=importHTML("http://en.wikipedia.org/wiki/China_at_the_Olympics","table",7)`
4. `=function("url", "object" , number)`

# Scraping data using Google spreadsheets and Import HTML

1. Open  
[http://en.wikipedia.org/wiki/China\\_at\\_the\\_Olympics](http://en.wikipedia.org/wiki/China_at_the_Olympics)
2. Open a Google Spreadsheet
3. Type in  
`=importHTML("http://en.wikipedia.org/wiki/China_at_the_Olympics","table",7)`
4. `=function("url", "object" , number)`

# Scraping data using Google spreadsheets and Import HTML

Webpage Screenshot

Untitled spreadsheet

File Edit View Insert Format Data Tools Add-ons Help

Comments Share

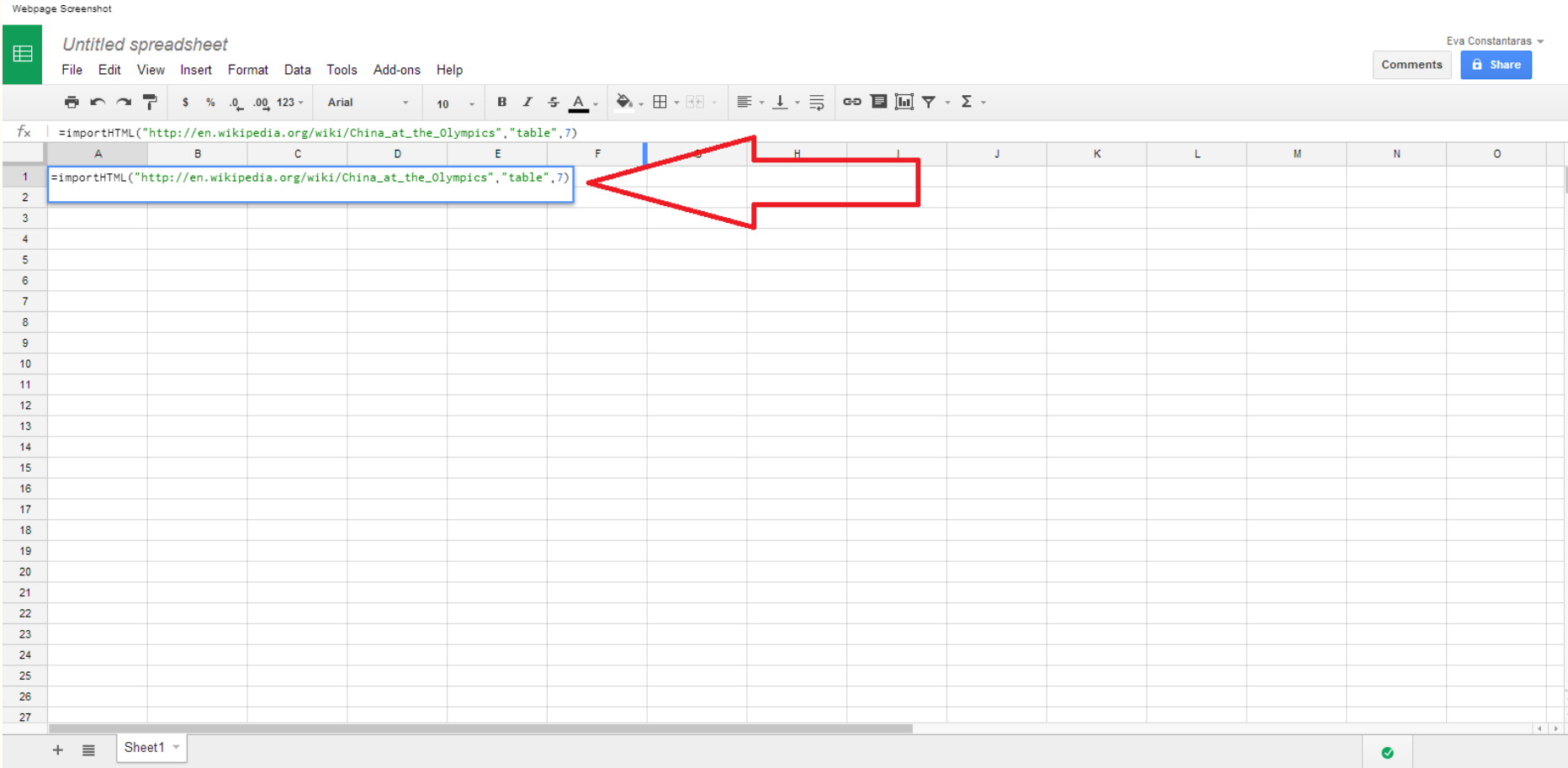
Eva Constantaras

fx =importHTML("http://en.wikipedia.org/wiki/China\_at\_the\_Olympics","table",7)

=importHTML("http://en.wikipedia.org/wiki/China\_at\_the\_Olympics","table",7)

Sheet1

https://docs.google.com/spreadsheets/d/1cWUcQ0eeTTCYc-QsH\_KekVdamdLbs-qskOgcnw1XHDw/edit#gid=0 Fri May 23 2014 10:50:29 GMT+0300 (E. Africa Standard Time)

The image shows a Google Spreadsheet interface. At the top, there's a menu bar with 'File', 'Edit', 'View', 'Insert', 'Format', 'Data', 'Tools', 'Add-ons', and 'Help'. Below the menu is a toolbar with various icons for formatting and editing. The spreadsheet itself has columns labeled A through O and rows numbered 1 through 27. In cell A1, the formula '=importHTML("http://en.wikipedia.org/wiki/China\_at\_the\_Olympics","table",7)' is entered. A red arrow points from the formula bar at the top to this cell. The formula bar also displays the same formula. At the bottom of the spreadsheet, there's a status bar showing the URL 'https://docs.google.com/spreadsheets/d/1cWUcQ0eeTTCYc-QsH\_KekVdamdLbs-qskOgcnw1XHDw/edit#gid=0' and the date/time 'Fri May 23 2014 10:50:29 GMT+0300 (E. Africa Standard Time)'. The spreadsheet is titled 'Untitled spreadsheet' and is owned by 'Eva Constantaras'.

# Contents

## Scraping data from the web

- Using browser plug-ins to scrape data tables
- Exercise
- Using Google docs to import data into charts
- **Exercise**

## Summary & Resources

# Scraping data exercise

1. Visit the [Mars Group budgets portal](#).
2. Find budget data that might enrich your reporting on tourism in Kenya
3. Try to scrape the budgets using both browser extensions and by importing from Google spreadsheets
4. Can you find a better way to search the Mars Group site? Try “filetype:” and “site:” shortcuts that you learned in the session on searching for data on the web.

# Summary and Resources

Need help scraping your data from the web?

- Ask the School of Data
- Join the data driven journalism listserve
- Join the Hacks/Hackers Nairobi Google group and go to the next meeting or post your question and ask for volunteers
- Visit the Internews data journalism team for support
- Register for a free import.io account and try it out



**Internews**  
Local voices. Global change.