

Introduction to Data Visualization Techniques Using World Clouds

Edited by Carolyn Talmadge Jonathan Gale, Revised by Kyle Monahan on April 17th 2017

Table of Contents

Introduction.....	1
Introduction to Word Clouds/Tag Clouds	1
When to Use a Word Cloud	2
How to create a Word Cloud using a generator on the web	2
Word Cloud Exercise.....	3
Resources	3

Introduction

This exercise provides an overview of basic best practices for text data visualization techniques using World Clouds. It covers determining when to use a word cloud and how to create word clouds using web-based tools.

Introduction to Word Clouds/Tag Clouds

Similar to a histogram, word clouds represent the frequency of certain words, terms or expressions. Unlike histograms, word clouds use **categories or terms from text**. They show the frequency of terms used in a paragraph, RSS feed, or other block of text by scaling the size of the terms and color to highlight the frequency of occurrence.

Unlike the other techniques for displaying information, word clouds are not easy to create in Excel; however, many web-based generators exist to create colorful and informative word clouds.

Figure 1 is a word cloud of the location of over 17,000 incidents in the *Armed Conflict Location and Event Data Project (ACLED)* database for 1997-2015 in Somalia⁷. This word cloud shows that the location with the most incidents over this time period was Mogadishu and scales all other incident locations to reflect their frequency in the database.

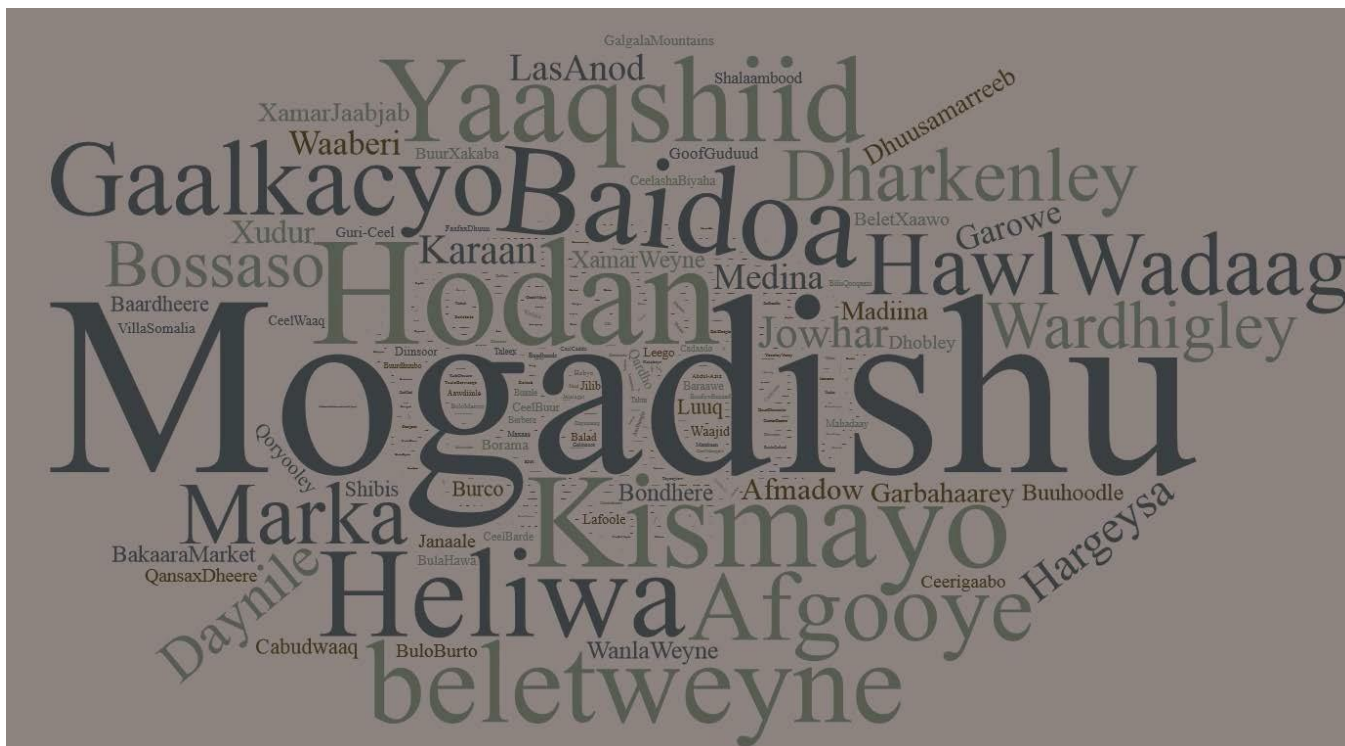


Figure 1: ACLED incident locations in Somalia 1997-2015.

When to Use a Word Cloud

Word clouds are useful to **qualitatively** display the frequency of many categories or terms within a large body of text or data. They are helpful when a dataset **has many categories or terms** beyond those that can easily be summarized in other charts or when using **unstructured data**. Unstructured Data refers to information that either does not have a pre-defined data model or is not organized in a pre-defined manner. Unstructured information is typically text-heavy, but may contain data such as dates, numbers, and facts as well. An example of unstructured data is a news article, report, journal article, etc.

How to create a Word Cloud using a generator on the web

Copy the desired spreadsheet column or document text.

1. Navigate to <http://timdream.org/wordcloud/>
2. In the popup, scroll down to the *Copy & Paste* option on the left-hand side and click it.
3. Paste the text into the text box that appears and click *Start* in the lower right corner.
4. Click the *Edit List* button and remove some commonly-used words in order to highlight more significant words or phrases. Click *Confirm* to regenerate the new word cloud.
5. Try exploring different styles using the various tools at the top.
6. The *Save* button allows one to save the word cloud graphic as PNG image file.
7. To create a new word cloud, use the *Back* button on the top left.

Word Cloud Exercise

Exercise data is located **S:\Tutorials & Tip Sheets\Tufts\Tutorial Data\Introduction to Data Visualization**

Use the instructions provided within the ***How to Create a Word Cloud Using a Generator on the Web*** section above to create the following word clouds from the different sources of information.

1. Create a word cloud from the PDF provided of the UN Secretary-General's Monthly Report, titled *s_2016_60.pdf* in the **Introduction to Data Visualization** folder. This document is about the Implementation of the Resolutions by All Parties to the Conflict in the Syrian Arab Republic⁸. Copy all the text from the PDF and paste it into the generator by highlighting all text (Control + A) and copying it (Control + C). This is an example of unstructured data visualized in a word cloud. Use the *Edit List* button to remove insignificant commonly used words to better highlight the meaningful terms. Try switching themes.
2. **No longer supported, for reference only:** With the same generator, create a word cloud using a *Reuters Africa* country level RSS news feed found here: <http://af.reuters.com/tools/rss>. Select a country. Then click the "View Feed XML" button. Copy and paste the new URL. In the word cloud generator site, click *RSS Feed*. Paste the RSS feed URL into the site and click *Start*. Use the *Edit List* button to remove "Reuters" and "Jan" and any other commonly used words to better highlight the meaningful terms.
3. Try using a global events web-based word cloud generator such as the *Global Database of Events, Language, and Tone (GDELT)*. Read an overview of *GDELT* here: <http://gdeltproject.org/#intro>
4. Next, create your town word cloud here: <http://analysis.gdeltproject.org/module-gkg-wordcloud.html>. You'll need to submit an email. It will take a few minutes to complete.

Resources

Web-based Word Cloud Generators

Timdream - <http://timdream.org/wordcloud>

Tagul - <https://tagul.com>