Creating Shaped Wordclouds Using R

Tidewater Big Data Enthusiasts
Chuck Cartledge
Developer

April 23, 2020 at 12:57 Noon

Contents

List of Figures

1 Introduction

The R library `wordcloud` provides an easy way to create an image showing how often a word (or tag) appears in a corpus (see Figure 1 on the following page). In a word cloud, the size of a word indicates how often that word appears. Word cloud words can be colored as well.

While word clouds are easy to create, often the clouds could be shaped differently to create a more lasting and profound impression (see Figure 2 on page 3).

2 Discussion

The R library `wordcloud2`[1] provides the capability of creating a word cloud that takes the shape of an image, or the shape of letters. The collection of predefined shapes include:

- 'cardiod' – a heart shape
- 'circle' – the default
- 'diamond' – an alias for a square
- 'pentagon' – the five sided object
- 'star' – a five pointed star
- 'triangle' – a triangle with the wide base at the bottom
- 'triangle-forward' – a triangle with the wide base at the left

This collection of shapes (when combined with a user specified background color), may be enough to satisfy a wide variety of needs. But it is the `figPath` option that offers the most potential.

The `figPath` option can point to a figure that contains the image the cloud path should fill.

Here are the steps to create an "interesting" shape to fill with a word cloud:

1. Download/create an image with only two items (see Figure 3 on page 4):
   - A white background, and
   - A black outline of the shape.
2. Fill the interior of the shape with the same color as the outline (see Figure 4 on page 5).
3. Pass the location of the filled image as the `figPath` parameter (see Figure 5 on page 6).

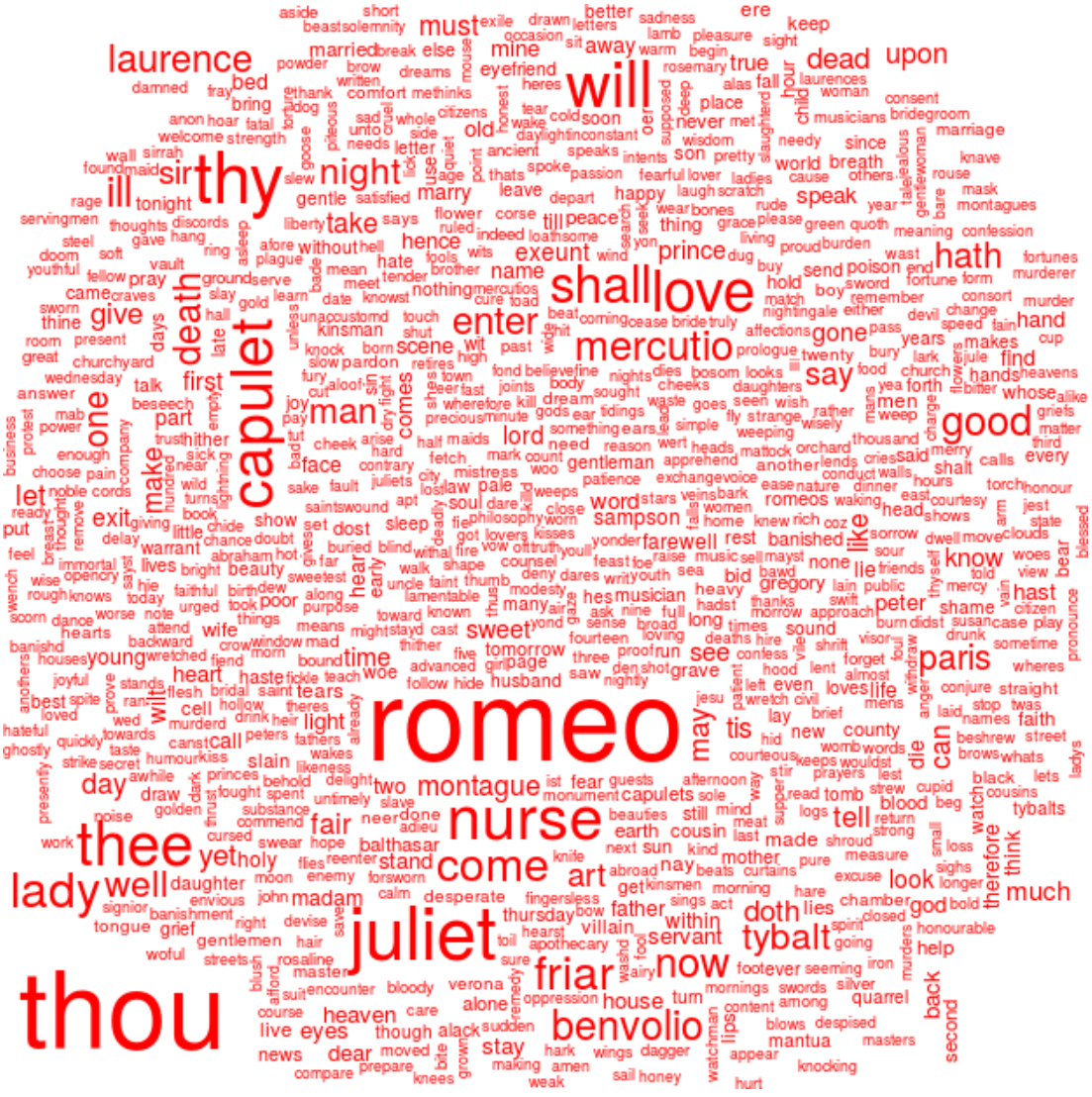


Figure 1: A sample word cloud based on Romeo and Juliet. The image was created using wordcloud function in the `wordcloud` library and the text from "Romeo and Juliet."

Figure 2: A more interesting word cloud based on Romeo and Juliette. The image was created using wordcloud2 function in the `wordcloud2` library and the text from "Romeo and Juliet."

Figure 3: An empty word cloud figure.

Figure 4: A filled word cloud figure.

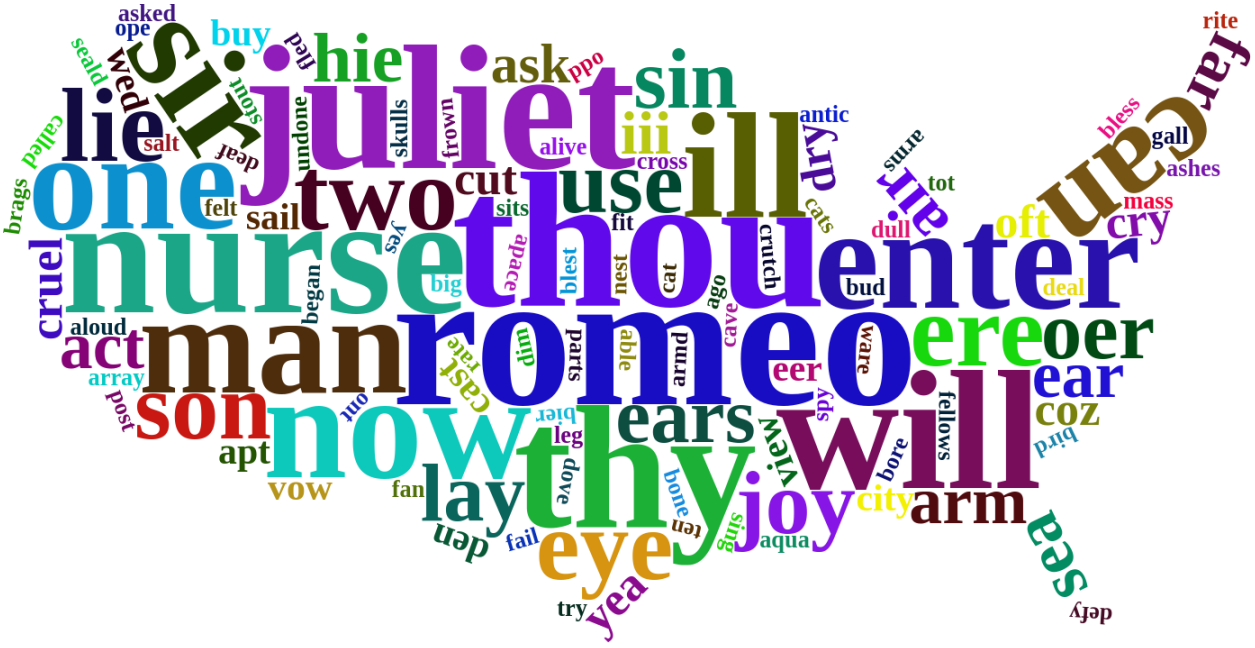


Figure 5: A filled USA word cloud figure.

The `wordcloud2` function behaves slightly differently than most of the other R plot functions that I've used. The result from both `wordcloud2` and `letterCloud` is not displayable within R. These functions actually create an HTML page in a temporary directory with embedded JavaScript that performs the placement of the words within the shape, and provides a level of interaction after the page is displayed. R "understands" that the product from these functions is an HTML widget and starts up the default browser to show the page. The page, and its sub-directories are removed when R ends.

The fact that the page uses JavaScript introduces some interesting aspects. Buried in the JavaScript used by the page to place the words in the cloud are a plethora of `Math.random()` calls. The JavaScript specification says that the `Math.random()` function has to return a value greater than or equal to 0, and less than 1, which is reasonable for a random function. The specification also says that the implementation of the random function is up to the JavaScript application, and does not specify how the numbers are to be generated. Meaning that the same HTML page being viewed by two different browsers, may generate two different sequences of random numbers. Most random number generators have the capability of setting a seed value so that a repeatable sequence can be generated. JavaScript does not support the idea of a random number seed. The HTML page and collection of directories can be moved to a server where they are available for use and support.

All of this means that each loading and viewing of the page will generate a different

[1] Available from:`https://github.com/Lchiffon/wordcloud2`

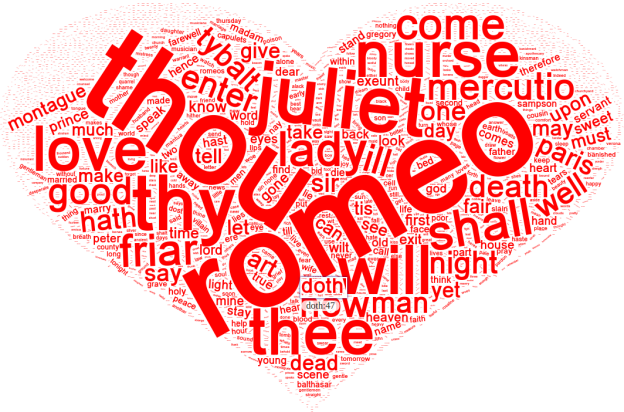
image, and there is no practical way to "get back" to an image that was good.

In the Files section (see Section A on page 9) is an R script and support files to work with. The R script was used to create various images (see Figure 6 on the following page).

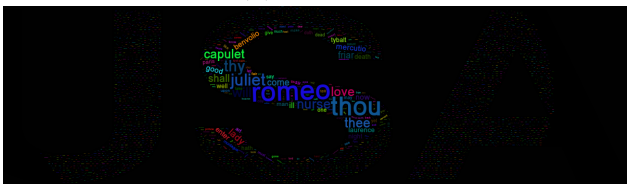
3 Conclusion

The `wordcloud2` library enables you to create word clouds of arbitrary shape inside an HTML using JavaScript to position and orient each word. Each HTML page and its associated library files are placed in individual directories that are removed when the creating R process terminates. Pages and files can be moved, or copied for safe keeping if desired. Because the pages use the Math.random() JavaScript function, each time the page is loaded, words will be positioned differently in the cloud. If the desired shape has an internal hole, then it is possible that some words may not be placed in the cloud.

`wordcloud2` allows you to create word clouds to support your data visualization needs.

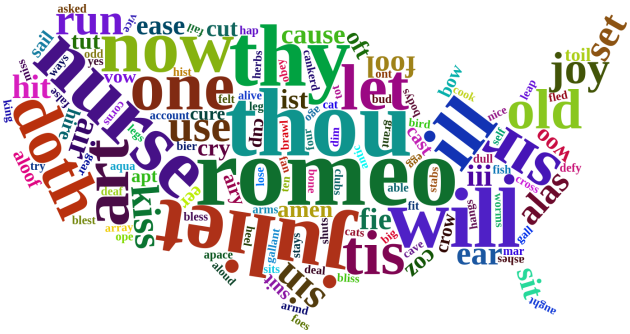


(a) A heart.



(b) The letters “USA”.

(c) A star.



(d) The USA.

Figure 6: A collection of sample word clouds. These images were created with the attached R script.

A Misc. files

The files used to create all these figures are attached to this report. They are:

1. romeoAndJuliet.base64 – default text used to demonstrate the software
2. heart.png – a heart shape with a hole
3. usa.png – an outline of the continental United States
4. wordCloud.R – an R script to demonstrate making word clouds