reimagining the word cloud

1) problem

Common text visualizations such as word clouds often leave out relevant context and remove dimensionality from the underlying data.

2) motivation

Text data, such as tweets, often correspond to a point in time and space that is not conveyed through traditional text visualizations. We encode these dimensions to provide underlying context.

3) approach

We use tweets during Hurricane Sandy, binned by distance (in miles) and time (in hours) from landfall in southern New Jersey on Oct. 29 at 8pm. Words are encoded with color based on their positivity, font size by relative frequency, and type-face by objectivity.

Possible extensions include a GUI allowing users to input their own text, using more refined part of speech tagging from Python’s Natural Language Toolkit, and encoding other dimensions that might result from text or literature.

Code @ github.com/dannycochran/WordRiver
Live @ stanford.edu/~c0chran/cgi-bin/WordRiver