



Fig. 5.8 An enhancement of the base event data figure that adds a color mapping for message keywords and pop-up ‘tooltips’ showing the full text of the twitter messages.

events – for example, suppose we want to explore a set of data on when people go to the movies in the United States. In 2008 well over a billion tickets were sold; plotting a symbol for each one would be tricky. Instead, we want to aggregate in some fashion. In fact, the data are reported already aggregated as weekly box-office totals. Even when aggregated to this level, we still arrive at a situation with nearly 7,000 movie/week combinations.

We could try a number of ways of representing this data. One possibility would be a 2D plot of *movie* by *time*; with dots at each intersection, *sized* or *colored* by the *attendance* variable. Unfortunately we have too many unique movies for this to be a useful technique. Worse, since each movie lasts only a certain amount of time, the display would be mainly empty, with all the interesting data compressed into small areas.

We solve this problem by using the *stack* position modifier again. In the same way that the Twitter chart (figure 5.8) stacked individual items, we can stack aggregates to equally good effect. A simple chart that would work in most applications is the stacked histogram – each week shows a *stack* with the movies for that week stacked on top of each other. If that chart is not available, then you may have to bin

bunched events need special techniques

