From the stacked bar graph of each year’s floods, we can see that year 2000-2010 has the most number of floods. Although the frequency of each year’s floods is different, the ratio of normal (M≤4 ), large (4<M≤6), and extreme (M>6) floods within each year is quite similar. Large floods appear the most, and then extreme floods, normal floods appear the least.

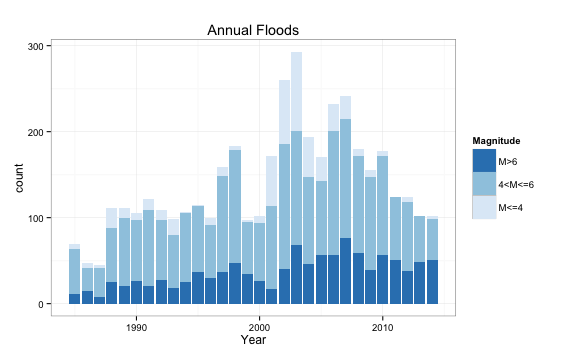


Figure <>: Stacked Bar Graph of Normal, Large, and Extreme Floods Each Year

We then aggregate the floods within each month and plotted the following stacked bar graph. Month 7 and 8 have the most number of floods, which makes sense because it is summer when there are a lot of rain.

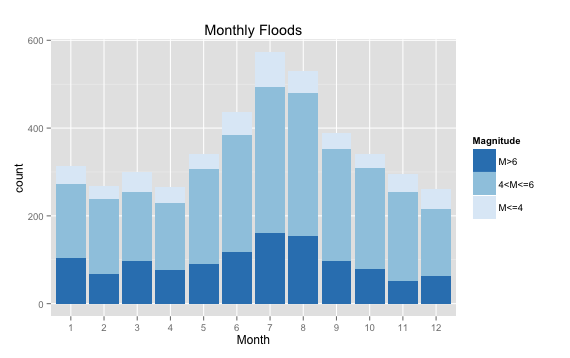
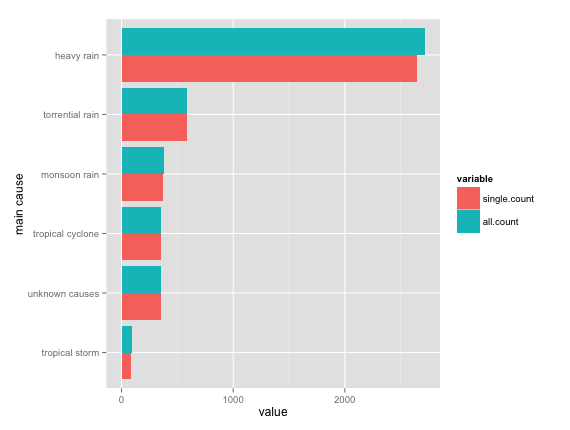


Figure <>: Stacked Bar Graph of Normal, Large, and Extreme Floods Each Month

We want to analyze the main causes of floods. Because heavy rain is the only overwhelming cause of floods whose frequency is more than four times of the second cause, torrential rain, we split the bar plots into two parts in order to see clearly how the less frequent ones distribute. There are two different groups of the bar plots, single count and all count. Single count is where we only count the floods which has that single cause and neglect the floods which have multiple causes, while all count is where we add the multiple causes that also contain that one cause. The top three causes of floods are all related with rain, which are heavy rain, torrential rain, and monsoon rain. We also notice that although dam break appears rarely in single cause, it has a much larger number in all count, which suggests that dam break is always combined with other causes to result in floods.



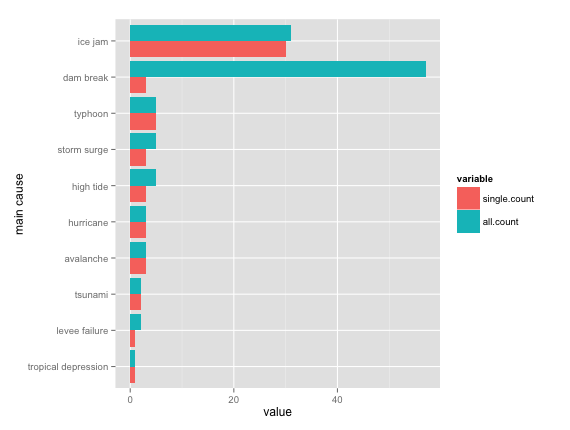


Figure <> Bar graph of main causes of floods (notice the two plots’ value is different)

We draw a word cloud of quoted headlines from copyrighted news stories to see which words are most concerned in the news. Obviously “flood” appears the most in the news, “water” ”rain” ”river” ”people” are also some major words.

