I retrieved my data from [www.usa.com](http://www.usa.com). Since it is in html format, it is difficult to gather data using R. Therefore I used bash to retrieve the data, during which I confronted with certain challenges. First the command “curl” doesn’t work in my notebook, so I used python to download the data, which is not so difficult. Second I reviewed regular expressions and used bash to “grep” those specific lines and then use “sed” to pipe all those data into a file. Third, by using Python, I managed to get rid of commas inside the data, and put all my data into dataframe.