*The following cleaning needs to be done:*

1. *For "State", removing characters from "[" to the end.*
2. *For "RegionName", when applicable, removing every character from " (" to the end.*
3. *Depending on how you read the data, you may need to remove newline character '\n'. '''*

These instructions transl

#algorithm or pesudo not real code

read the file as lines

lst=[]

state=''

region=''

for each line

if line contains '[ed' then the line has a state

state = get the state by removing everything from '[' to

                end

else

if the line has a '('

region = get region by removing everything from '(' to end

else

region = line with \n removed

endif

add state,region to lst

endif

endfor

This process can be done either through a for loop or through apply e.g. this [stackoverflow](http://stackoverflow.com/questions/41457322/pandas-rearranging-a-data-frame/41458629" \l "41458629" \t "_blank) article, note the example handles only cases with "(" and "[" so an additional if statement is needed in the regionName lambda





