All code, files reside here:
https://github.com/karunmj/usu-coursework/tree/master/cs5660datasc/hw/hw4

This is a solution for task 2 regarding counties that committed fraud. I have written a python script to process the reducer function results (mapper and reducer functions were mentioned as appendix in main document).

Task 2: Election fraud

3. Counties where voter fraud has occurred in 2008.
   These 15 counties have a percentage increase in number of votes per party greater than 50%.

County ID	Party ID
178	3
241	3
334	3
244	3
332	3
274	3
390	3
220	3
359	3
201	3
303	3
424	3
324	3
107	3
474	3

Appendix: Fraud pre-processing python script

import pandas as pd

fraud = pd.read_csv("outputel37.txt", header = None, index_col = False, sep = "\t")

countyid = fraud[0]

count2006 = fraud[1]
count2006 = count2006.map(eval).apply(pd.Series)

count2008 = fraud[3]
count2008 = count2008.map(eval).apply(pd.Series)

#Increase in 2008 for party 1

inc1 = ((count2008[1]-count2006[1])/count2006[1])*100

#Increase in 2008 for party 2
inc2 = ((count2008[2]-count2006[2])/count2006[2])*100

#Increase in 2008 for party 3
inc3 = ((count2008[3]-count2006[3])/count2006[3])*100

fraudproc = pd.concat([countyid, inc1, inc2, inc3], join='outer', axis=1)

fraudproc.to_csv(path_or_buf="fraudproc.txt")

sort1 = fraudproc.sort_values([1], ascending = 0) #None above
sort2 = fraudproc.sort_values([2], ascending = 0) #None above 50
sort3 = fraudproc.sort_values([3], ascending = 0) #Couple above 50