Description of data:

You will find three datasets, which are from the City and County Data Book, composed of 1994 and 1990 demographic data on all counties in the U.S. along with presidential voting tendencies for the 1992 U.S. presidential election. These datasets consist of 16 variables. You will report all your results in a PDF file generated via RMarkdown, with your code embedded into your results and discussion. Perform the following analysis on this data:

1. Combine these datasets into one data file and perform the following analysis on the data using R. and discuss the method you used to combine these datasets into your report.

2. Calculate the percentage missing in each of the variables in this data, also discuss various solutions to deal with missing data. Use the one you think is appropriate in this situation.

3. Identify if there are any outliers in the data? If any exists then discuss the possible solutions to deal with outliers, suggest and apply the possible solution in this situation.

4. Write down a function that will help calculate five values summaries (min, max, mean, median, SD) for all the continuous variables present in the data. Use this function to calculate 5 values summary for each continuous variable and report your results in the form of a table (generated by your function) and also report it in your RMarkdown report.

5. Develop R Shiny flexdashboard to report your statistical analysis, a. Construct a table that shows the average with Standard deviation of population size, population density, number of democrats and number of republicans, number white and black population and turnout from each state. b. Construct a bar plot that displays the comparison of average black and white population in each state.

6. Discuss what is parallel programming and how it can be performed in R. Use parallel programming to run your function to perform Exploratory Data Analysis on the given dataset. Report how parallel programming helped speed up your code.