Stat 133, Spr '08
Primary Election Project
Part II: Due Friday May 16, midnight
Part III: Due Tuesday May 20, midnight

For the second part of this assignment, you will conduct most of your analysis. Your analysis must include the following three components:

Recursive partitioning It is expected that you use recursive partitioning to fit a predictor for whether a county will favor Clinton or Bush in the 2008 Democratic primary. Play around with the arguments to the *rpart* function in R to settle on a particular classification tree. Continue your analysis to compare your partition to the one presented in the NY Times (see the figure at the end of this assignment). Dig around a bit deeper to see if there are alternative trees that will do just as well.

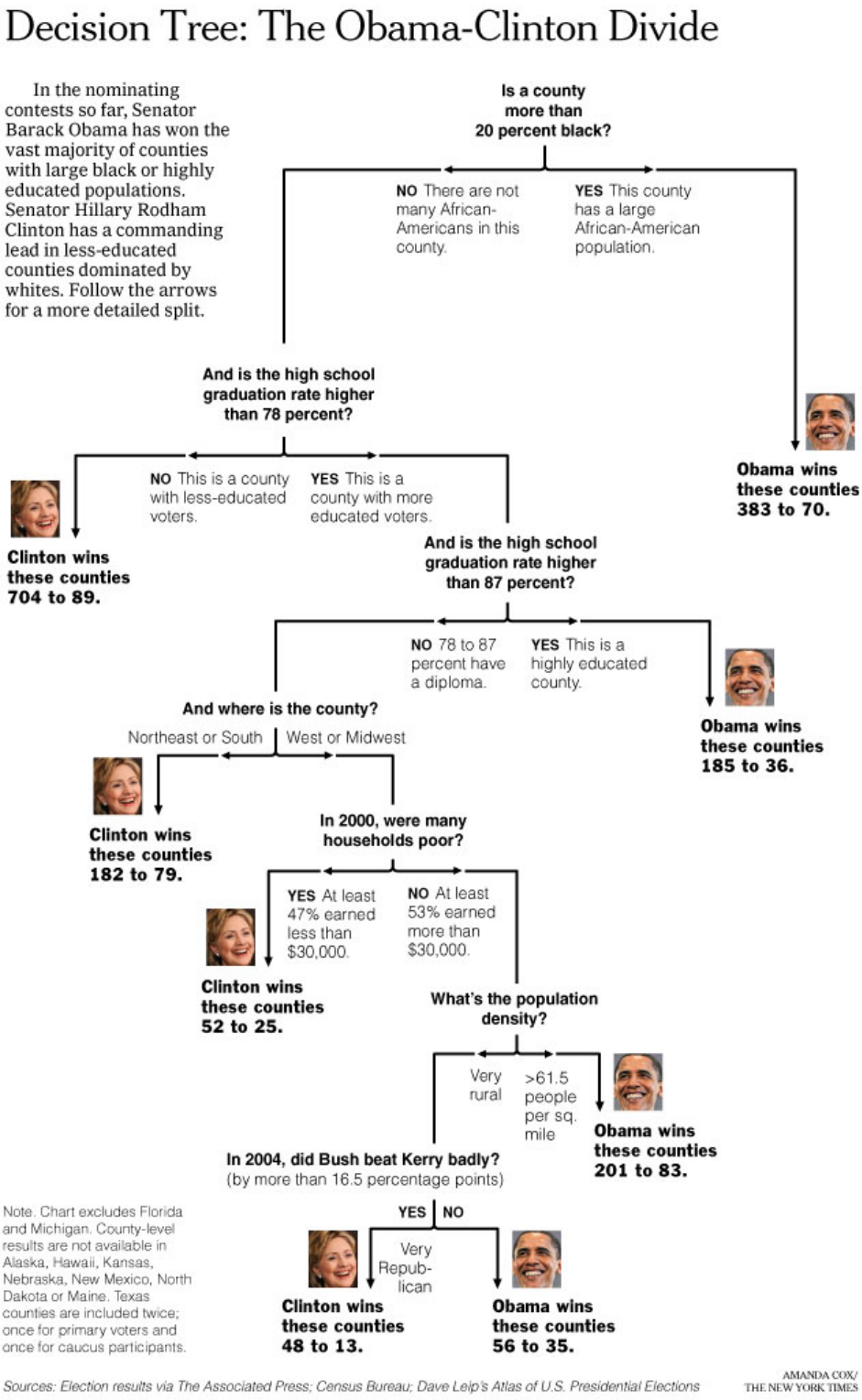
Map Use the *maps* library to create a map that brings out the geographic nature of these data. Where are the Clinton supporters located? Try including additional data on your map, such as the census data. Use as your guide the 2004 election map that appeared in the NY Times as shown in the figure at the end of this assignment.

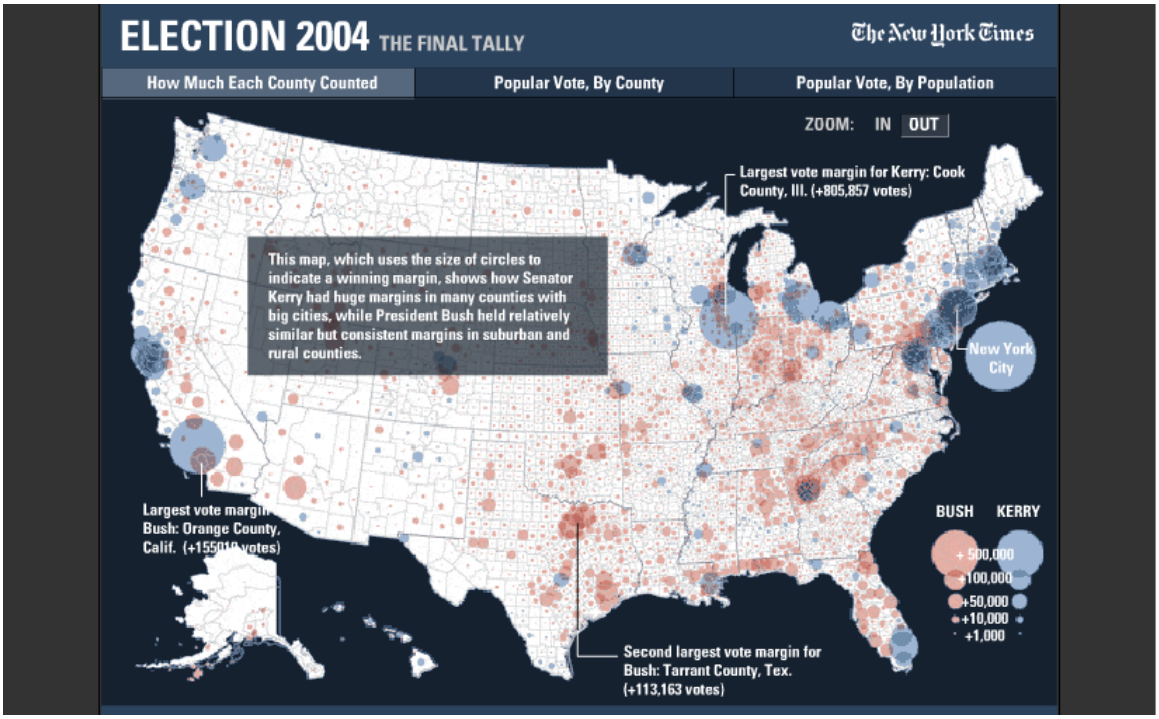
Visualizations Continue to examine your data for other interesting features related to the county characteristics and the outcome of the Democratic primary.

For Part II You are to provide the code in a plain text file, and the images that you are planning to use in your final write-up. My goal is to provide feedback to you about your plots before you write up your final report.

For Part III You are to provide your final write-up, including the write-up from the first part of the assignment (you are welcome to revise for submission in this part). The report

must be no longer than 10 pages, including plots. Make sure that your plots are well made. Use the guidelines from the elephant seal project for guidance on how to write your report.




ELECTION 2004 THE FINAL TALLY
The New York Times
How Much Each County Counted
Popular Vote, By County
Popular Vote, By Population
ZOOM: IN OUT
Largest vote margin for Kerry: Cook County, Ill. (+805,857 votes)
This map, which uses the size of circles to indicate a winning margin, shows how Senator Kerry had huge margins in many counties with big cities, while President Bush held relatively similar but consistent margins in suburban and rural counties.
New York City
Largest vote margin for Bush: Orange County, Calif. (+155019 votes)
Second largest vote margin for Bush: Tarrant County, Tex. (+113,163 votes)
BUSH KERRY
+500,000
+100,000
+50,000
+10,000
+1,000