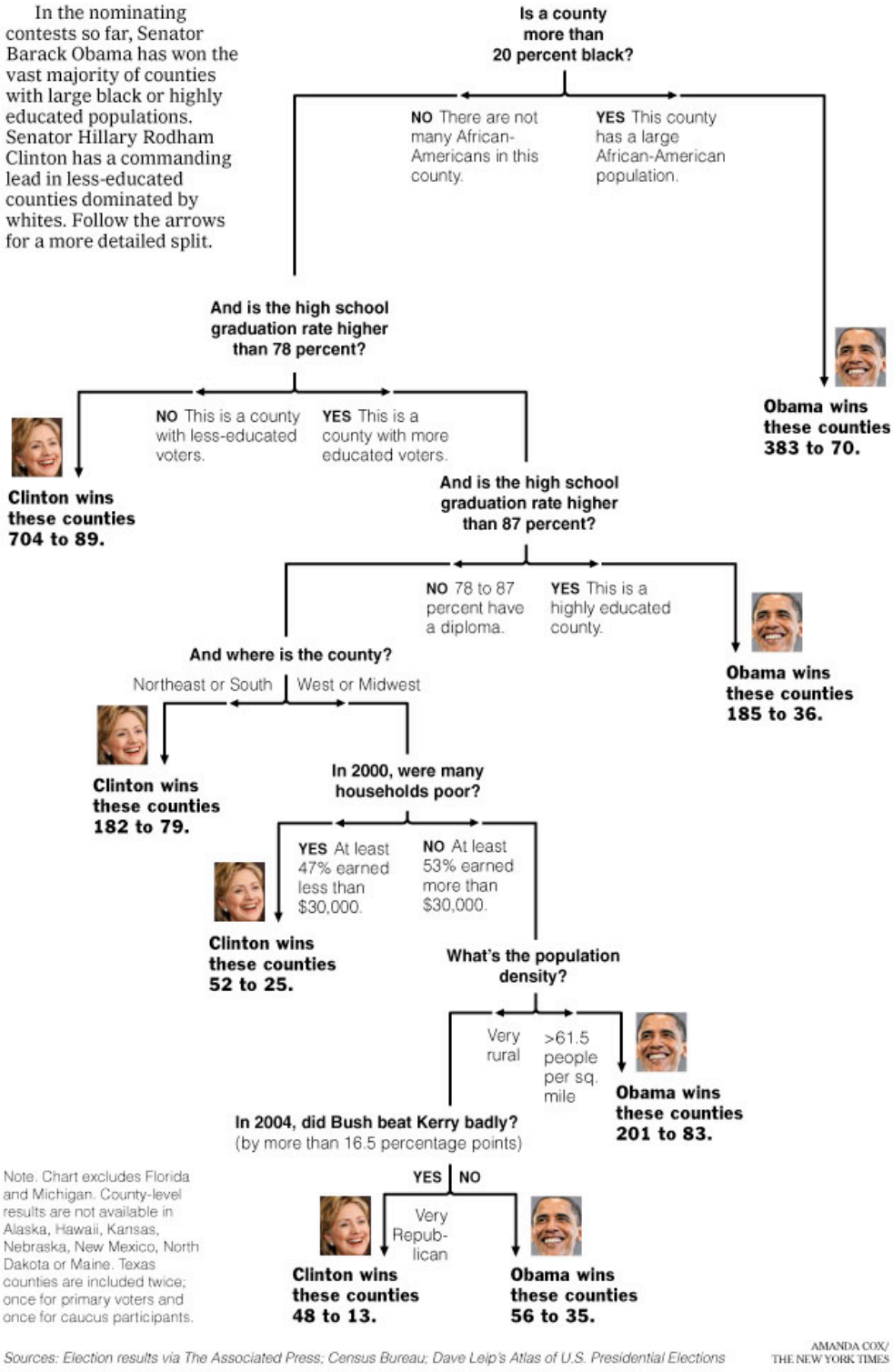
2008 Democratic Primary

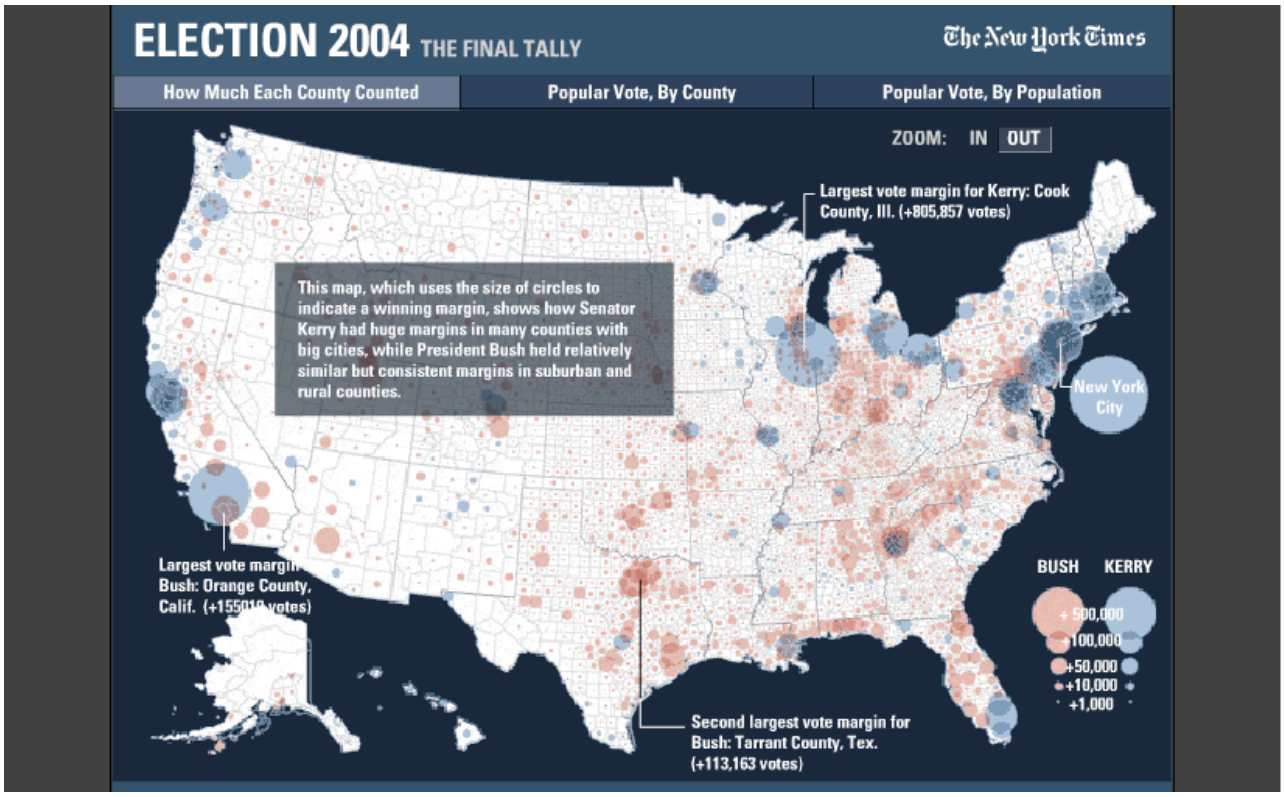
Which Counties have Clinton/Obama wins?

New York Times, Wed Apr 16, 2008

Decision Tree: The Obama-Clinton Divide



Are there regional trends?



Questions

- Are there differences between the counties that go for Clinton and those that go for Obama?
- Where are the Clinton-counties located?
- How well can you predict the primary results for the upcoming primaries?

Data

- 2000 Census data at the county level: Excel files
  - cc07_tabB1.xls: population, land area, and density
  - cc07_tabB3.xls: race, age, gender
  - cc07_tabB4.xls: HS degree, Bachelor's degree, foreign born population, persons in poverty, HH income >$75,000
- 2004 Presidential election: countyVotes2004.txt scraped from CNN website
- 2008 Democratic Primary: DemPrimary2008.txt scraped from CNN Website
- Geographic data: counties.gml latitude and longitude
- Auxiliary data:
  - top 30 cities in country
  - mapping of states to regions
  - mapping of townships to counties for New England States

Background

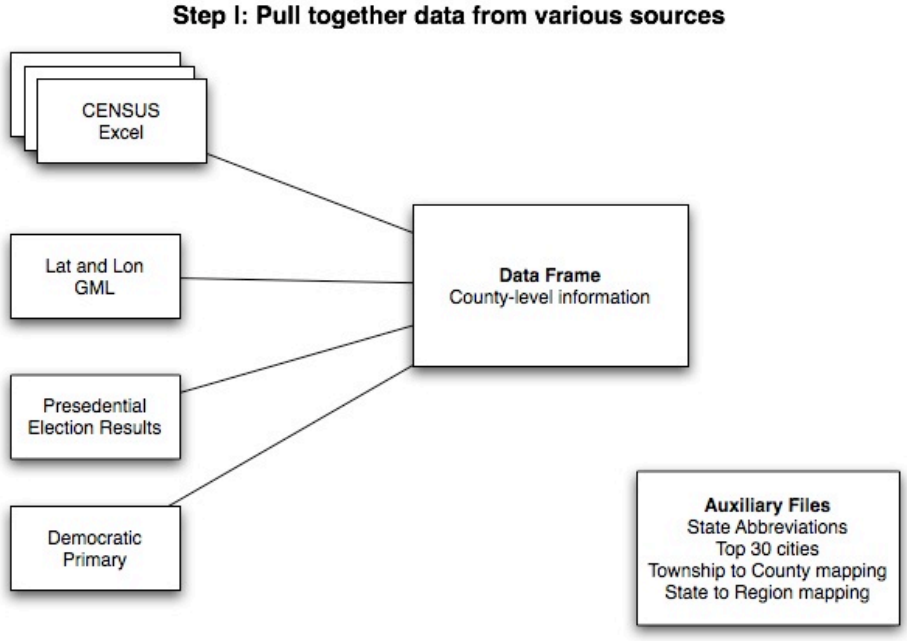
The CNN site presentation of the data changed from 2004 to 2008.

- In 2004, it was available in HTML tables, and could be acquired through programmatic calls to the Wed and post-processing the HTML.
- In 2008, the tables were generated on the fly in javascript and json. We used a packet sniffer to figure out the correct URL to call. Then we programmatically called the URL, received json in return, and processed it with RJson package to get plain text.

Administrative Details

- Final project
- Intermediate deadlines over 3 weeks
- Work in groups of 4-5
- Write up; 8-10 pages, plus 6 plots, and code

Step I: Data preparation

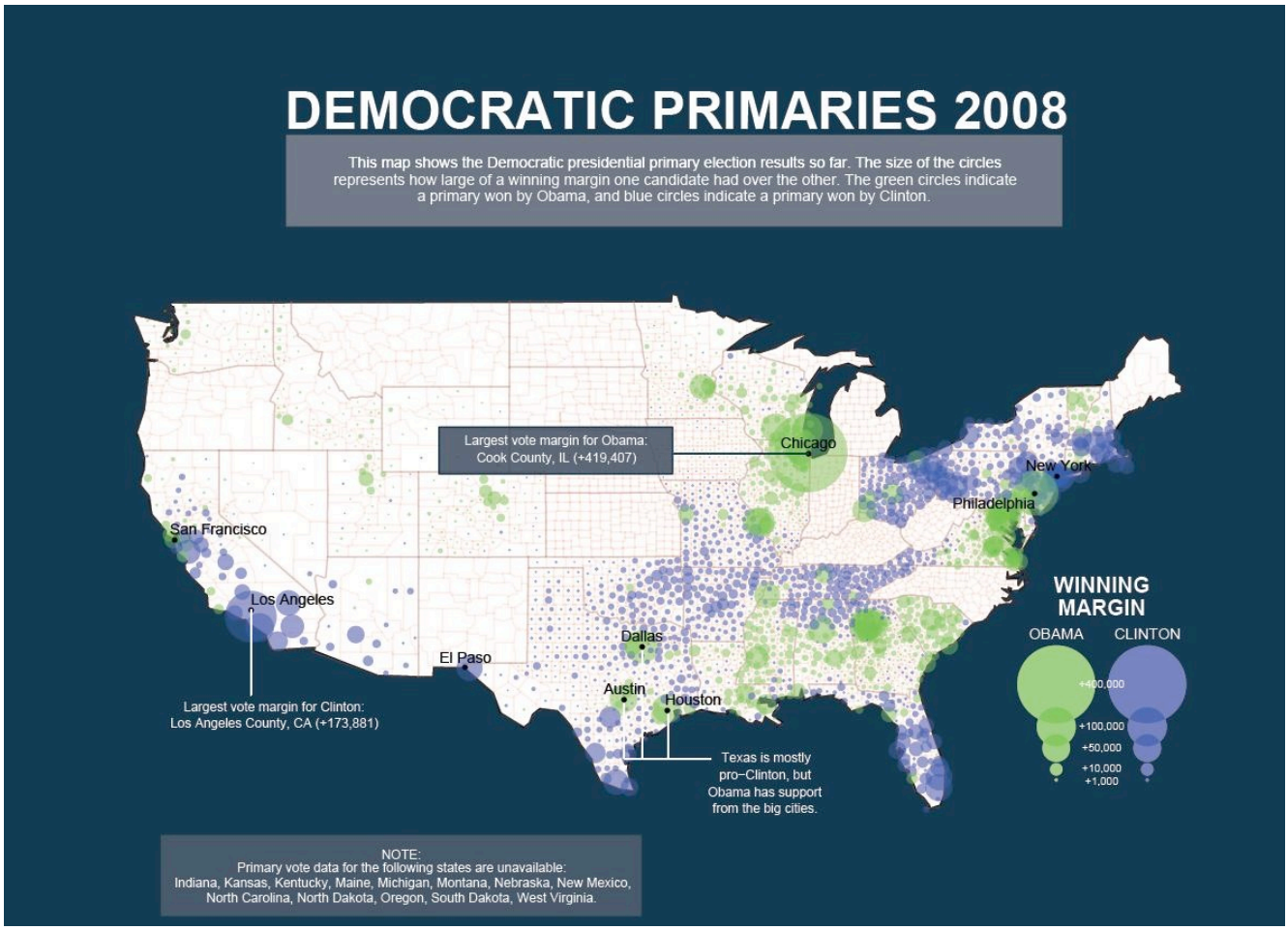


Use EDA to confirm that the data have been correctly extracted and merged.

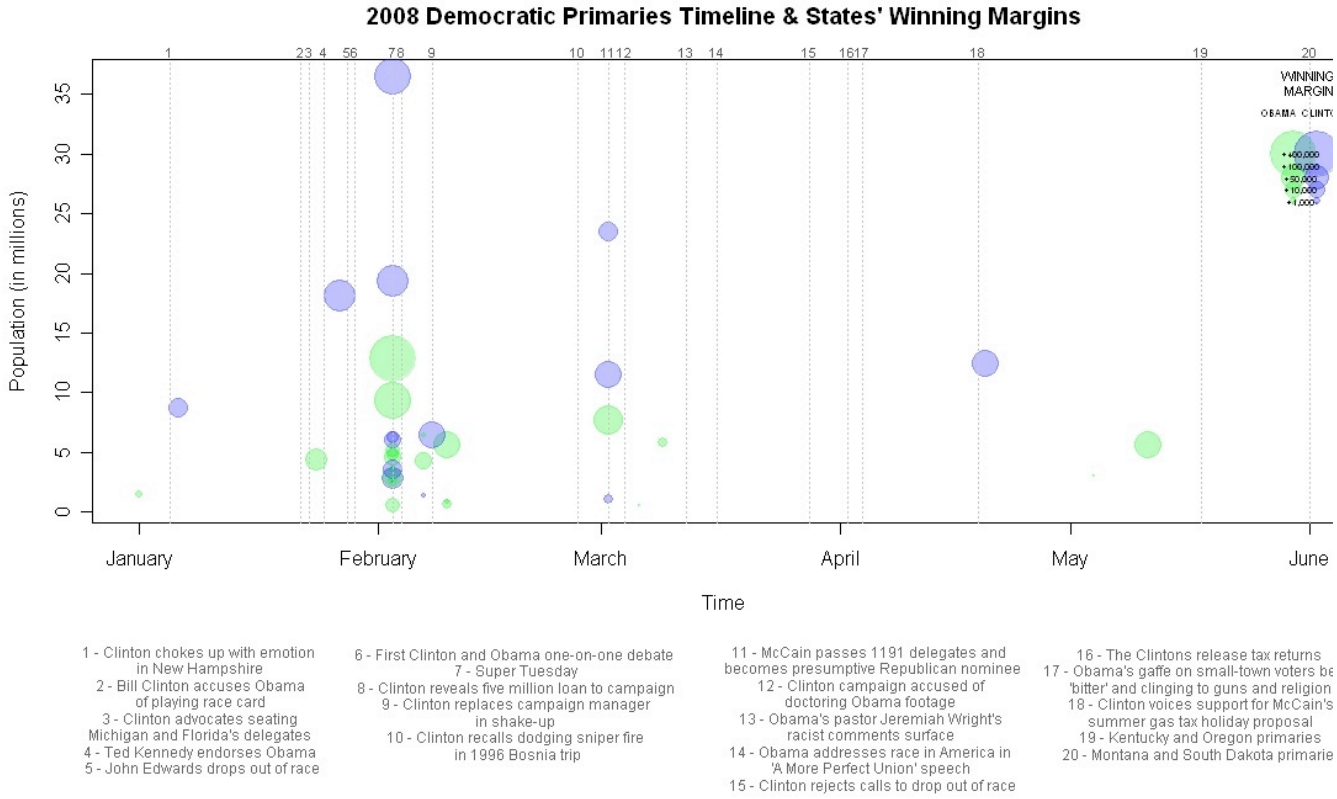
Step II: Model Fitting and Map making

Provide plots for comment.

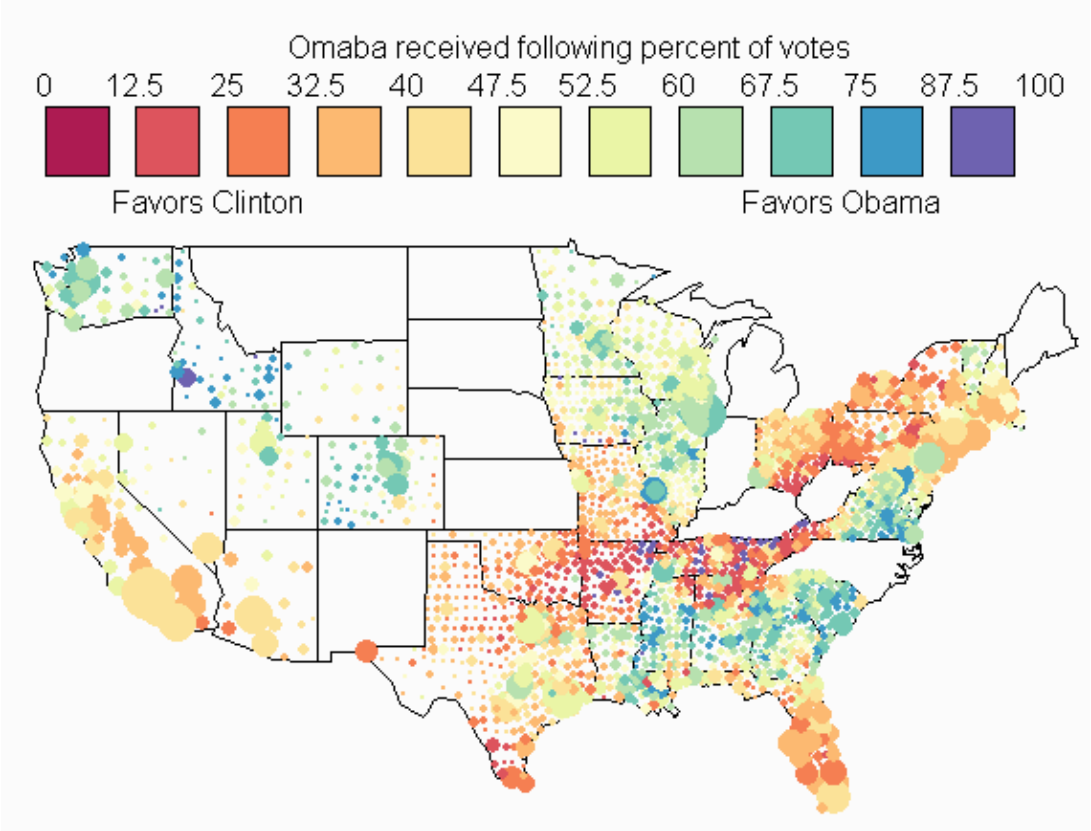
This group of students followed the color schemes of the NYT, and they added markers that make it clear that Obama tends to get support in the big cities.



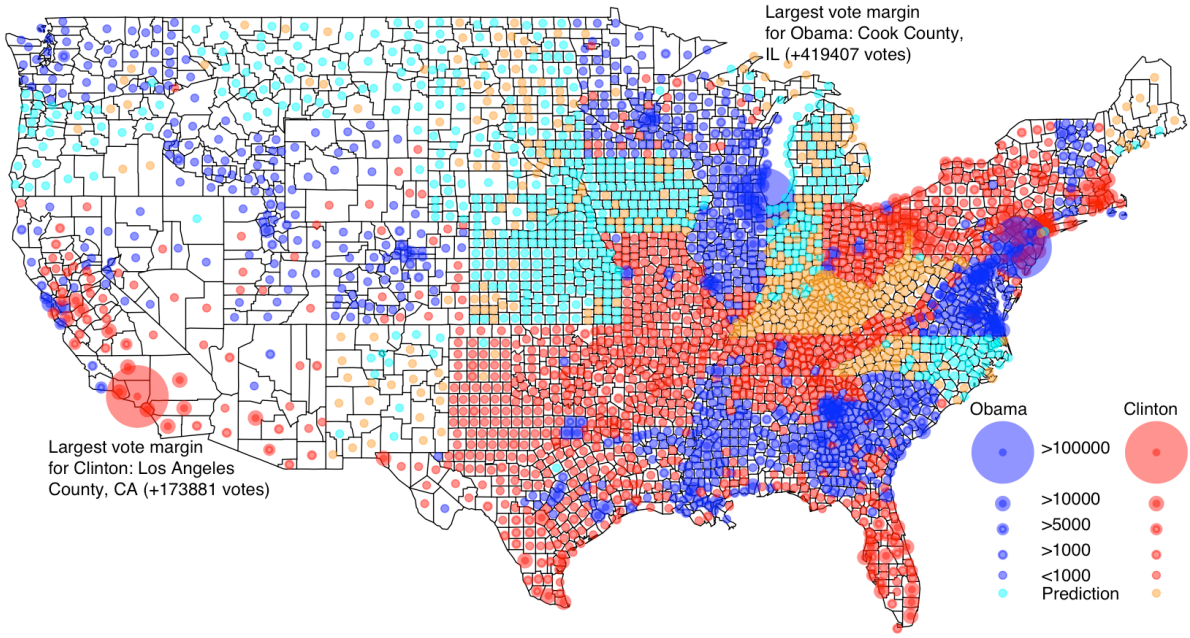
This group of students examined how the state primaries unfolded in time. The size of the circle denotes the vote margin, the color denotes the winner, and there are many markers to indicate political events.



These students display the percentage victory through color, in addition to showing the size of the victory through area of the circle.



These students used their recursive partitioning results to predict the Obama/Clinton wins and place the predictions on the map. It shows the Obama advantage in the deep south, the Clinton advantage in the Appalachians, and the close race in the midwest.



Step III: Final Write-up

Students turn in, report with graphics embedded, appendix of how the data frame was created, and code. Samples will be posted on the web.

Last modified: Fri Jul 11 17:00:47 PDT 2008