

Code books for ga_ballots.dat and ga_registration.dat
Spring 2006

The files are available in the following Athena locker, or via download from the subject web site:

/mit/17.871/Examples

ga_ballots.txt is a tab-delimited file that contains information about ballots cast in the 2004 presidential election in each county in Georgia. The following are the variables:

County FIPS code (“FIPS” stands for “federal information processing standard” and is a unique identification number for each geographic entity in the United States, used across all federally-produced data sets.)

County name

Number of absentee ballots cast in the 2004 election

Total number of ballots cast in the 2004 election

Total number of “provisional ballots” cast in the 2004 election (A provisional ballot is cast when the voter goes to the polls and isn’t on the registration list. This allows the voter to cast a ballot and then for the registration confusion to be cleared up later.)

Total number of provisional ballots counted in the 2004 election.

ga_registration.txt is a fixed-field file that contains demographic information about registered voters in Georgia counties in the 2004 presidential election. The numbers in parentheses indicate how many columns wide the variable is.

County FIPS code (5)

County name (8, string)

Number of black females registered (6)

Number of black males registered (5)

Number of white females registered (6)

Number of white males registered (6)

Number of Asian-American females registered (4)

Number of Asian-American males registered (4)

Number of Hispanic females registered (4)

Number of Hispanic males registered (4)

Number of others registered (5)

Total number of registrants (6)

Do the following:

(1) Save the ga_ballots and ga_registration raw data files as a Stata file in your Athena locker.

(2) Merge the ga_ballots and ga_registration data files by county.

(3) Generate two new variables: (a) percentage of ballots cast as provisional ballots and (b) percentage of registrants who are black.

(4) Generate a scatterplot with these two newly-generated variables, with (a) on the y-axis and (b) on the x-axis.