Data Source:

Two primary data sources are used to extracting and aggregating population/voting population in each county and each congressional district. Both of them are obtained from the Census.

1. Block Population: This data is based on the “2010 Census [P.L. 94-171] Summary Files " with the link <http://www.census.gov/rdo/data/2010_census.html>. Its technical document is available at <http://www.census.gov/prod/cen2010/doc/pl94-171.pdf>. All data are only available at the state level, which could be downloaded from the census website or FTP. I use the latter and all data are saved in the folder “R:\CDCongressionalCrosswalk\Census\_Block\_Population”.
2. 113th Congressional District Crosswalk Table: This data is from the “Block Equivalence Files” on the webpage “113th Congressional and New State Legislative District Plans” with the link <http://www.census.gov/rdo/data/113th_congressional_and_new_state_legislative_district_plans.html>. This file contains each state’s all blocks (which ID is unique across the whole nation) and corresponding congressional district ID. It also contains a national file which has aggregated all states’ information.

Data Manipulation:

There are two options to extract data from original files with PL extension: (1) Using Access to create expected dataset for each state; and (2) Using STATA to extract them. Both of them require the conversion of the original file from a PL file to a txt file through modifying the extension of each file. The instruction of how to use Access to extract data from the file is available through the link <http://www.census.gov/rdo/pdf/0HowToUseMicrosoftAccessShells.pdf>. However, this method requires an one-by-one operation for each state. After all targeted Access files are obtained, we can follow the information shown on the webpage <http://www.ats.ucla.edu/stat/stata/faq/odbc.htm> to transfer the Access file to dta file and then aggregate them. The STATA do file which contains codes of directly reading targeted variables from txt files and aggregating them is available at "R:\CDCongressionalCrosswalk\Census\_Block\_Population\block\_cd\_pop\_comments.do”.