

TEXAS SLAVERY PROJECT

Statistical Methodology for Estimating Missing Population Data

In this project we predicted missing values for the population of slaveholders and slaves from the counties of Texas from 1837 to 1845. Using visualization techniques, we grouped counties with similar population patterns and trends, using the arc of population shifts in particular counties to estimate changes in other similar counties whose data was missing. The population trends in some of the counties seemed to have followed trends that differed from most counties, and in those cases we used the average of the preceding and exceeding years to estimate the missing data.

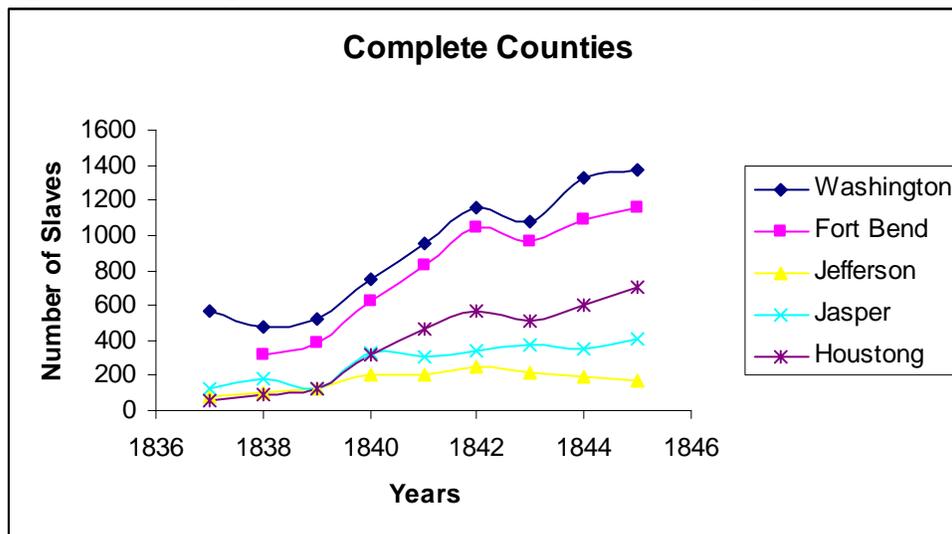
Data

The data consists of 30 different Texas counties that existed between 1837 and 1845, with nine sets of variables for each county. The first of these variables is the year for which the population size of various group has been gathered from original sources. The remaining eight variables represent different groups of the county's population. The first two of these groups are the total number of slaves and total number of slaveholders. The remaining of the variables partition the slaveholder group into subgroups based on the number of the slaves held (1-4 slaves, 5-9 slaves, 10-19 slaves, 24-49 slaves, and 50 or more slaves). The last variable represents the number of slaves owned by the slaveholder who had the largest number of slaves within that county in a particular year. While some of these counties existed throughout the entire period, others came into existence in later years as counties sub-divided into new counties, meaning that for some counties we had nine observations of the data while in others we had fewer.

Examples of Methodology

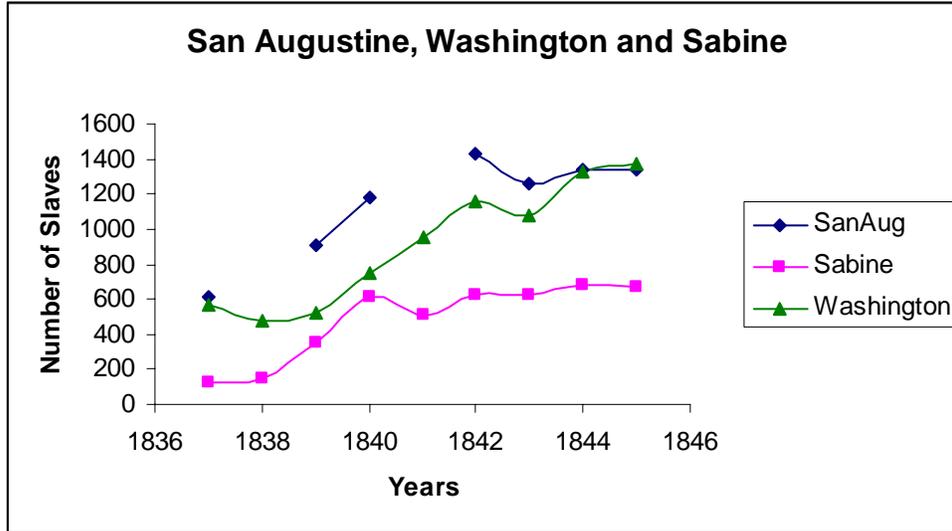
Figure 1 depicts some of counties for which we were able to obtain complete data for the years 1837 to 1845. We used these trends as well as the patterns for other complete counties as a guideline for estimating the missing points for the remaining counties.

Figure 1: Slave Population trends of some counties with complete data



From Figure 1, it is apparent that populations in some counties followed very similar trends: for example Washington and Fort Bent counties follow a parallel trend from 1838 to 1845, and for 1837 we do not have to estimate the population size for Fort Bent. We can use Washington County, therefore, as a base for further predictions of Fort Bend County's 1837 population.

Figure 2: Slave Population trends for San Augustine, Washington and Sabine Counties

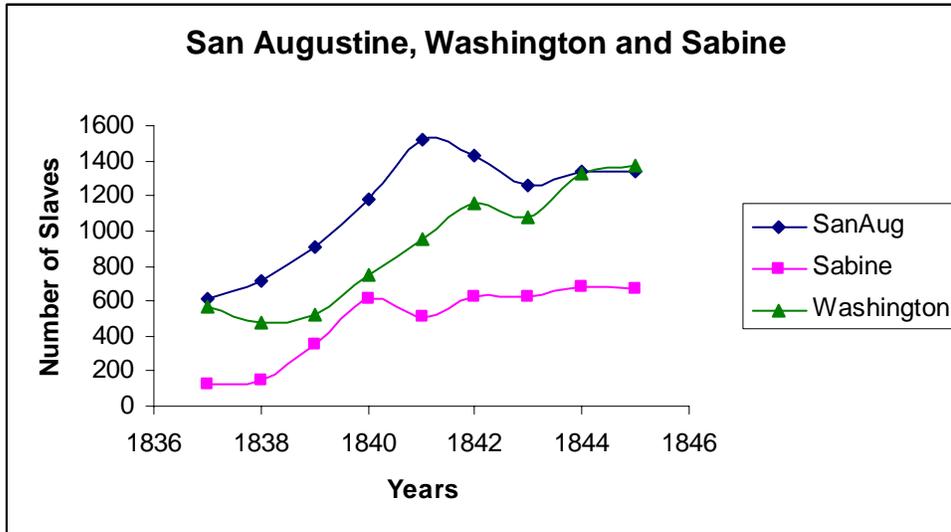


To estimate the missing values for the slave population of San Augustine County, we plotted the interrupted trend of this county along with Sabine and Washington counties on Figure 2. It appears that the population changes in San Augustine County followed a similar trend to Sabine and Washington counties. We therefore used the Sabine County data to guide the estimation of the first missing observation of San Augustine County. Because the 1840-1845 population trend of the San Augustine County resembles the population trend of Washington County from the same period, we used the population arc in Washington County to estimate the second missing point of San Augustine County.

The number of slaves in Sabine County in 1837 and in 1838 was 128 and 148 respectively. The number of slaves, in other words, increased 1.156 times or by 15.6 percent. In 1837 the number of slaves in San Augustine County was 614. Assuming that the number of slaves had grown at similar rates in these two counties, the estimated number of slaves for San Augustine County in 1838 would be $614 \times 1.156 = 710$. Thus we estimated the missing slave population for this year as 710. To fill the second gap of the San Augustine County which corresponds to 1841, we calculated the rate of change of the slave population of the Washington County from 1840 to 1841. This rate was $957/744 = 1.2829$. That is to say, there was 28.29 percent increase in slave population in Washington County. Therefore, we estimated that in 1841 the number of slaves in San Augustine County was $1185 \times 1.2829 = 1520$.

These techniques produced the estimated trend of the San Augustine County displayed in Figure 3, along with the trends of Sabine and Washington counties.

Figure 3: Estimated Slave Population of San Augustine along with Washington and Sabine counties



We used similar analysis to fill in the missing points for the remaining counties. In some cases, however, we were unable to find similarities in population trends. For those cases, we used the average of the preceding and succeeding observations of those counties to estimate the missing values. And if the missing data corresponded to the first year, we used the succeeding two points of that county to estimate the missing data.

The same analysis techniques were used to estimate the missing data points for the slaveholder populations. The change in the total population of slaves in a county was used to estimate the holdings that year of the county's largest slaveholder. If the county's slave population was estimated to have gone up by 23.45 percent in that year, for example, the holdings of the county's largest slaveholder was estimated to also have gone up by 23.45 percent.

In counties where these estimations were calculated, the various sub-categories of slaveholders (such as those who held 1-4 slaves, 5-9 slaves, etc.) may not always sum up to the total given for estimated total slaveholder population.