

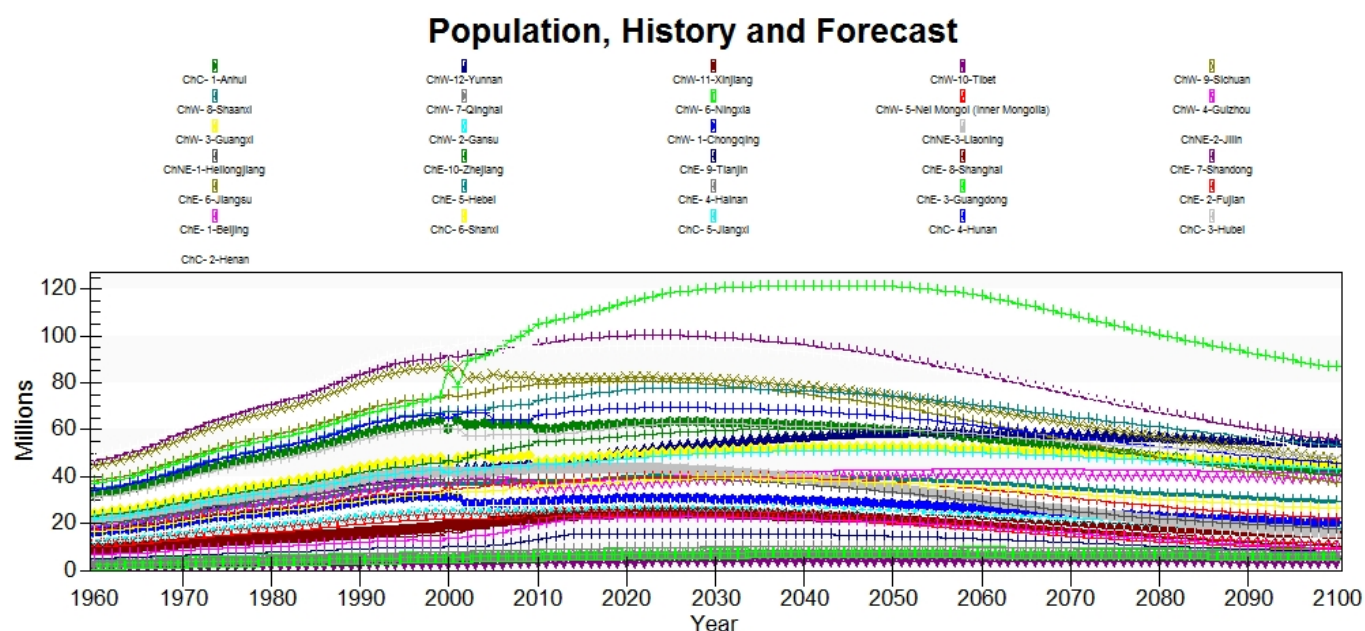
# Alternative Data Sources

The printable version is no longer supported and may have rendering errors. Please update your browser bookmarks and please use the default browser print function instead.

## China Statistical Yearbook

This provincial population series for China runs from 1995 through 2015. The original data lacked observations for Chongqing in 1995 and 1996 because Chongqing did not gain its status as a municipality until 1997. Chongqing was part of Sichuan province prior to 1997 and thus, Sichuan province's data included Chongqing in 1995 and 1996 and there was a substantial drop in Sichuan province in 1997. To remedy this issue, the ratio of population in Chongqing relative to the sum of Sichuan province and Chongqing in 1997 was used to estimate the population in Chongqing in 1995 and 1996. This population estimate for Chongqing was subtracted from Sichuan's population in 1995 and 1996. This provides a smoother and more accurate historical series. and data for Chongqing and Sichuan province were adjusted/estimated for 1995 and 1996.

Despite this estimation, there are still some spikes and drops in the population data around 2000. Guangdong has the most noticeable shift, where in 1999 population is 73.9 million and in 2000 population jumps up to 86.9. These shifts in the data are indeed in the data and not human errors. The data source was contact regarding this and other potential data issues that were found while pulling data, they have yet to respond. It is believed that the jumps in the data can be explained through redistricting/rezoning.



This provincial population series came from the China Statistical Yearbooks from 1996-2016. These yearbooks are available online on the National Bureau of Statistics of China website. The population data is published in tens of thousands, rather than the millions that are used in the full 186 version of IFs. Rather than converting the population data into millions, the population series was imported as it was published in tens of

thousands and ApplyMultAll is selected in the data dictionary. This normalizes the historical data to the population of China in the full 186 version of IFs. This choice to normalize rather than change units is meant to decrease the likelihood of human error, and simplify future data updates.

## Central Statistics Organization

In previous versions of the China subregional model a population series that ran from 1960 through 2013 was used. The previous data dictionary entry for this series said that the population series was sourced from the Central Statistics Organization and that there was data for 1999 and 2003, which was clearly incorrect. The source has not been found and some of the data was questionable. The prominent red flag was the appearance of what may have been sorting errors, where data for some provinces appeared to be switched over several years, and some unexpected observations for Chongqing. There is a single data point in 1982 that is as high as the population throughout the 2000s. Moreover, the next observation is the mid 1990s is about a tenth of the 1982 value. After 1998 Chongqing's population rapidly increases ten-fold to around 30 million. All of this made this unknown population source unusable.

## China Data Center

The China Data Center database from the University of Michigan published population data from 1949 through 2003. This data was considered as a potential source to blend with the current population data from the China Statistical Yearbooks to provide a longer historical time series. However, the two series could not be blended because the data from 1995 through 2003, where the two series had overlap, did not match.

---

Retrieved from "[https://pardeewiki.du.edu//index.php?title=Alternative\\_Data\\_Sources&oldid=3574](https://pardeewiki.du.edu//index.php?title=Alternative_Data_Sources&oldid=3574)"

---

This page was last edited on 25 May 2017, at 21:42.