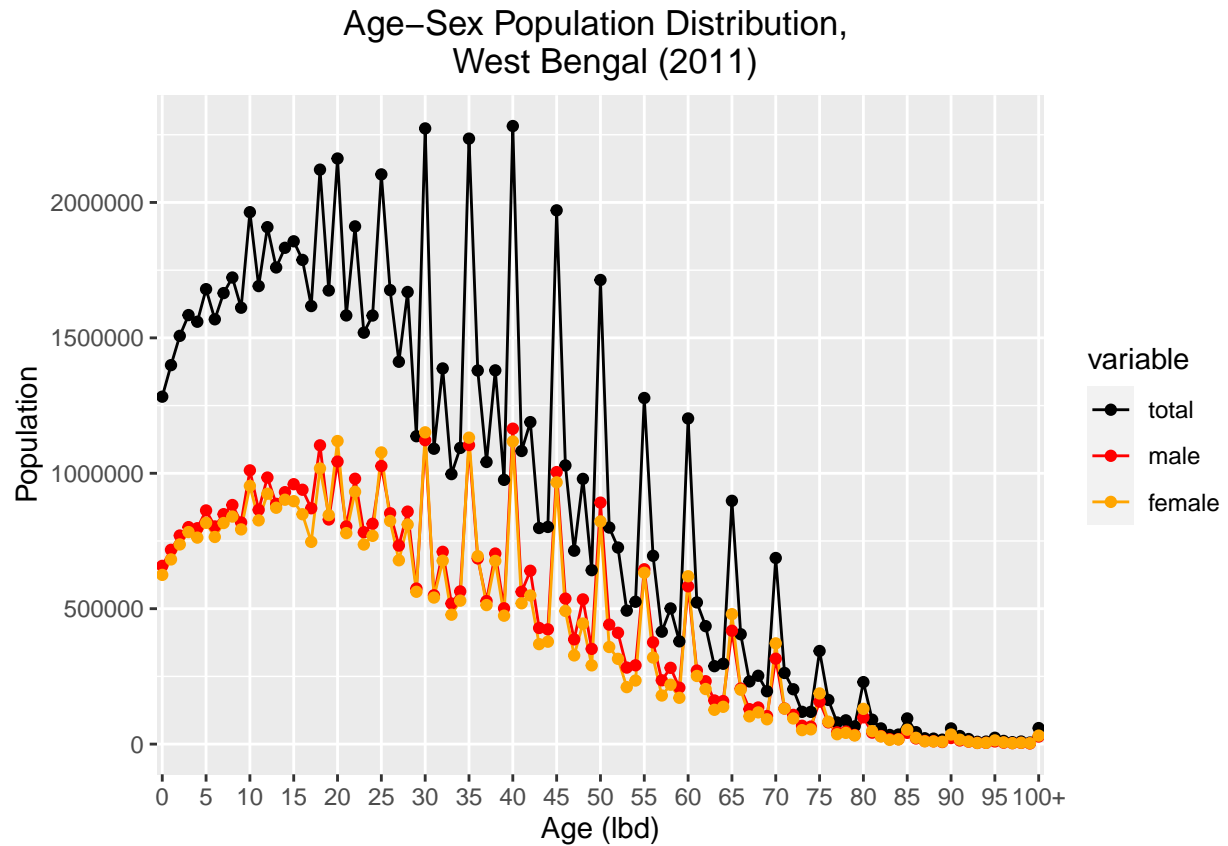# Error in age-sex Data of WB

Saikat Roy
IIPS/MBD/2020-22/28

23/12/2020

The errors in the age-sex data of West Bengal as enumerated in the census of West Bengal, 2011 are discussed here.
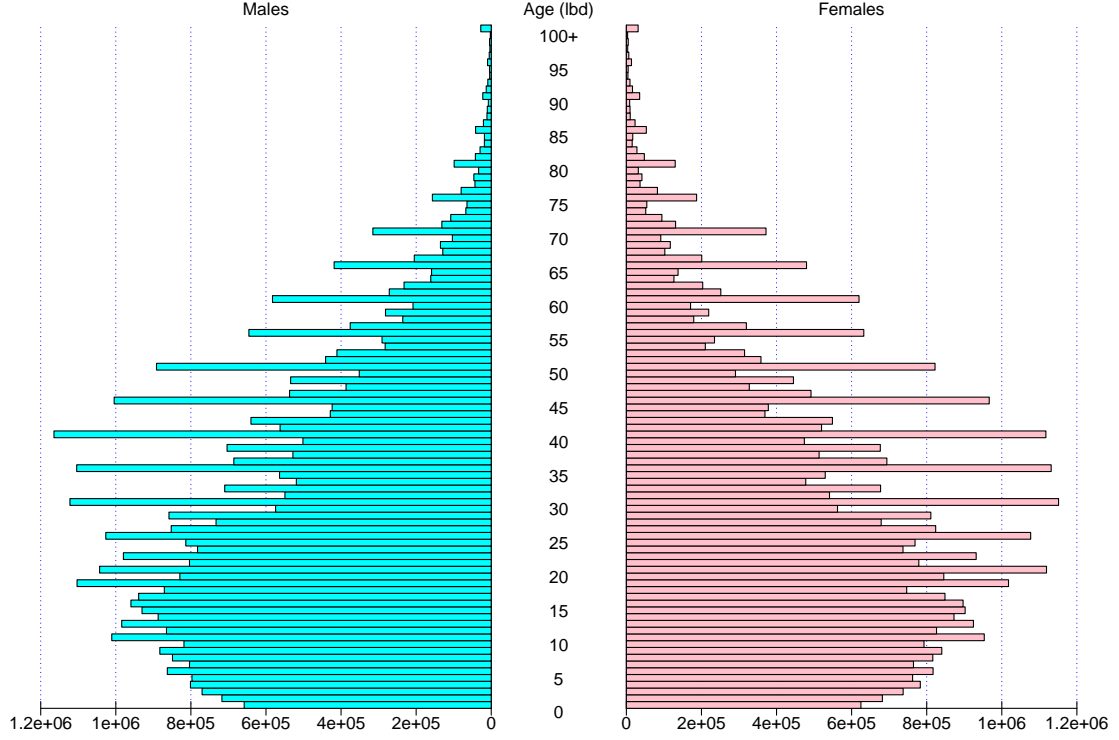
Single-year Age Groups

Line curves are plotted for total, male and female populations according as the ages.

Single year population pyramid with absolute population size values for each age-sex is constructed.

Single–year Age Population Pyramid
West Bengal, 2011

It is clear from the above figures that population count unusually spikes hugely at ages ending 0 and 5. Spikes, albeit smaller, are also noticed at even ages. In reality, the age-sex distribution ought to be smooth and devoid of all the spikes. It is clear that people are biased in favour of reporting ages divisible by 5 and slightly lesser biased towards reporting even ages. The tendency to report odd ages is low. This age-heaping is a prime source of errors.

The **Whipple's index** measures the extent of preference for ages ending with 0 and 5 under the assumption of rectangularity or linearity assumption in a ten-year age range.

Given that P_x is the population of West Bengal of the cohort with lbd x, the *Whipple's index* for heaping on terminal age digit '0' is measured as,

$$WI_0 = \frac{P_{30} + P_{40} + P_{50} + P_{60}}{\dfrac{P_{23} + P_{24} + P_{25} + \ldots + P_{62}}{10}} \times 100 \; = \; 161.97$$

and that for heaping on terminal age digit '5' is measured as,

$$WI_5 = \frac{P_{25} + P_{30} + P_{35} + P_{40} + P_{45} + P_{50} + P_{55} + P_{60}}{\dfrac{P_{23} + P_{24} + P_{25} + \ldots + P_{62}}{5}} \times 100 \; = \; 163.23$$

The *Whipple's index* for both ages ending 0 and 5 are very high. This denotes that the data is not accurate due to high digit preferences for 0 and 5 which results in age-heaping.