Regional workshop on the Production of Population Estimates
and Demographic Indicators
Addis Ababa, 5-9 October

Evaluation and Analysis of Age and Sex Structure

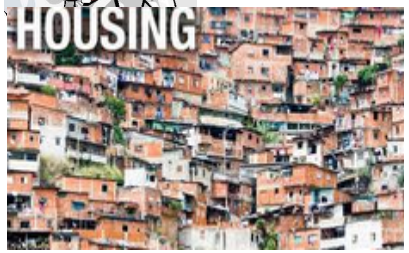
François Pelletier & Thomas Spoorenberg
Population Estimates and Projections Section

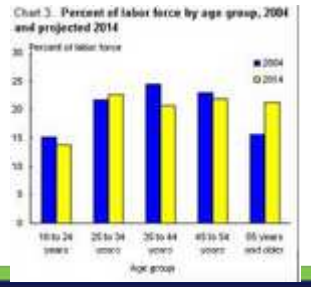
Evaluation method of age and sex distribution data

- Basic graphical tools
  - Graphical analysis
    - Population pyramids
    - Graphical cohort analysis
  - Age and sex ratios
  - Summary indices of error in age-sex data
    - Whipple's index
    - Myers' Blended Method

Importance of age-sex structure

- Planning purposes – health services, education programs, transportation, labour supply
- Social science, economist, gender studies
- Studying population dynamics – fertility, mortality, migration
- Insight on quality of census enumeration
- Having strong effect on other characteristics of a population
  - Determined by fertility, mortality and migration, and follows fairly recognizable patterns





What to look for at the evaluation

- Possible data errors in the age-sex structure, including
  - Age misreporting (age heaping and/or age exaggeration)
  - Coverage errors – net underenumeration (by age or sex)

- Significant discrepancies in age-sex structure due to extraordinary events
  - High migration, war, famine, HIV/AIDS epidemic etc.

Approaches to collecting age and its impact on quality

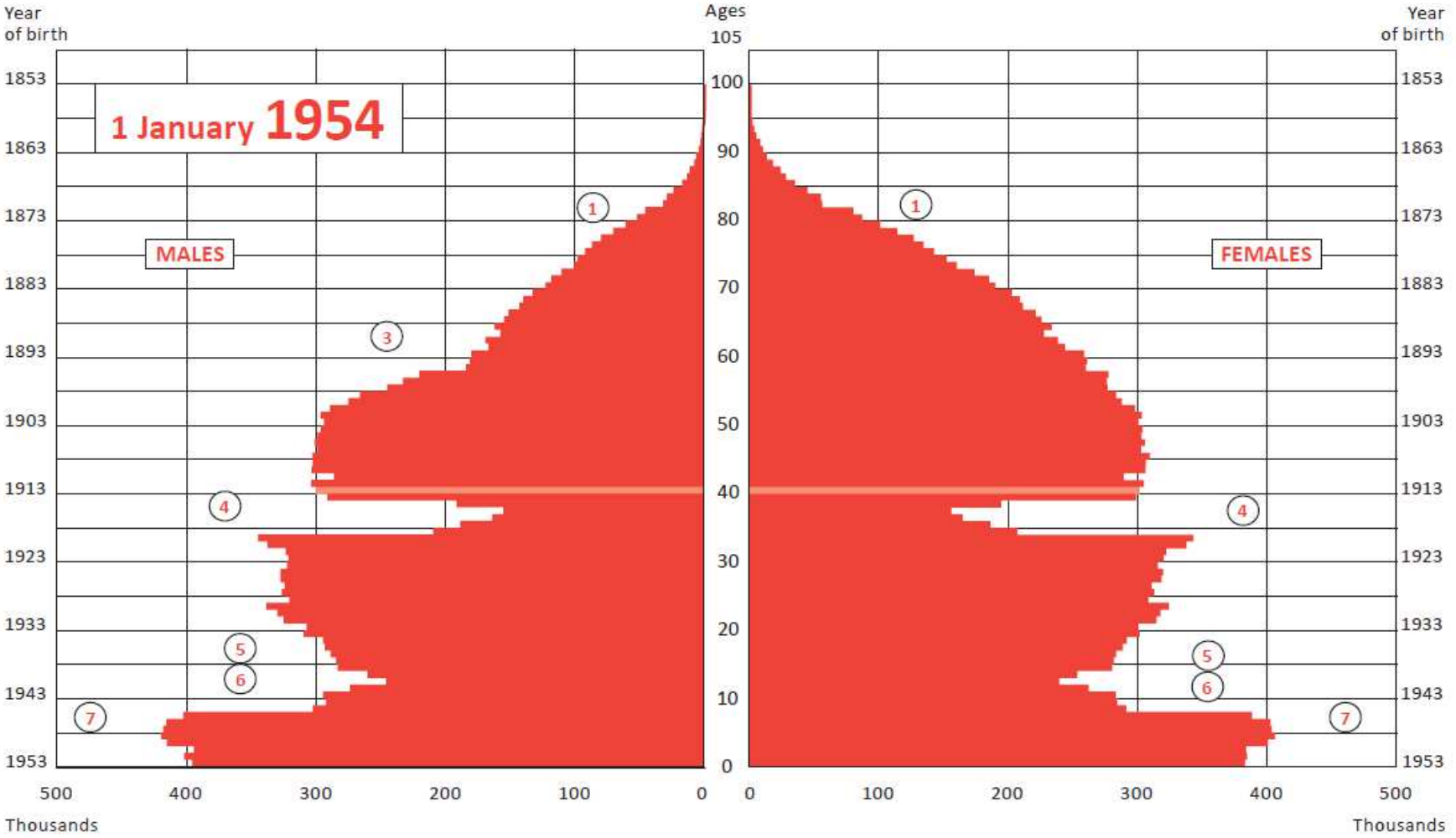
- Age - the interval of time between the date of birth and the date of the census, expressed in completed solar years

- Two approaches
  - The date of birth (year, month and day) - more precise information and is preferred
  - Completed age (age at the individual's last birthday) – less accurate
    - Misunderstanding: the last, the next or the nearest birthday?
    - Rounding to nearest age ending in 0 or 5 (age heaping)
    - Children under 1 may be reported as 1 year of age

Basic graphical methods – Population Pyramid

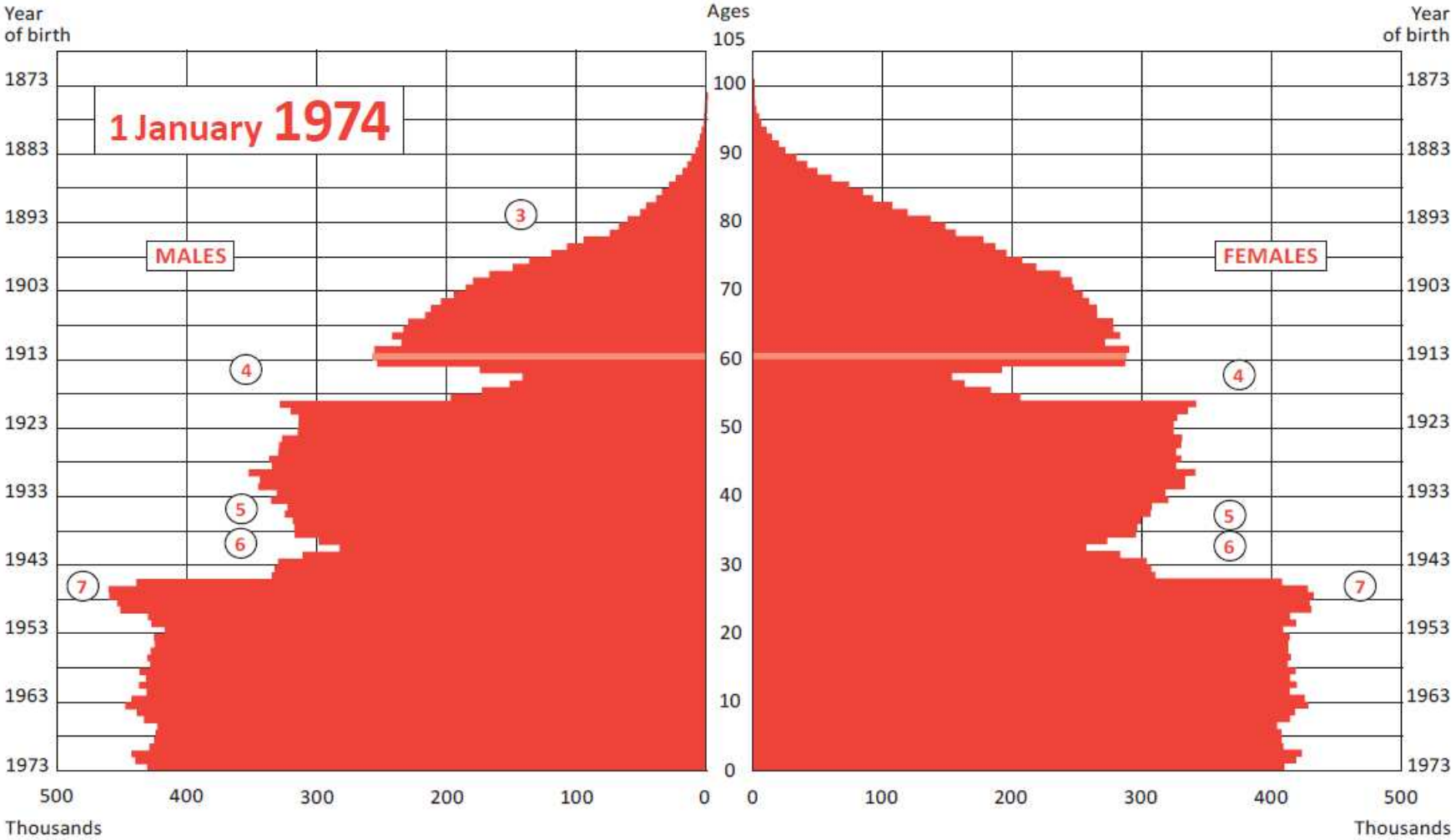
- Basic procedure for assessing the quality of census data on age and sex
- Displays the size of population enumerated in each age group (or cohort) by sex
- The base of the pyramid is mainly determined by the level of fertility in the population, while how fast it converges to peak is determined by previous levels of mortality and fertility
- The levels of migration by age and sex also affect the shape of the pyramid

Population pyramid – France, 1954 (*Source*: Pison 2014)



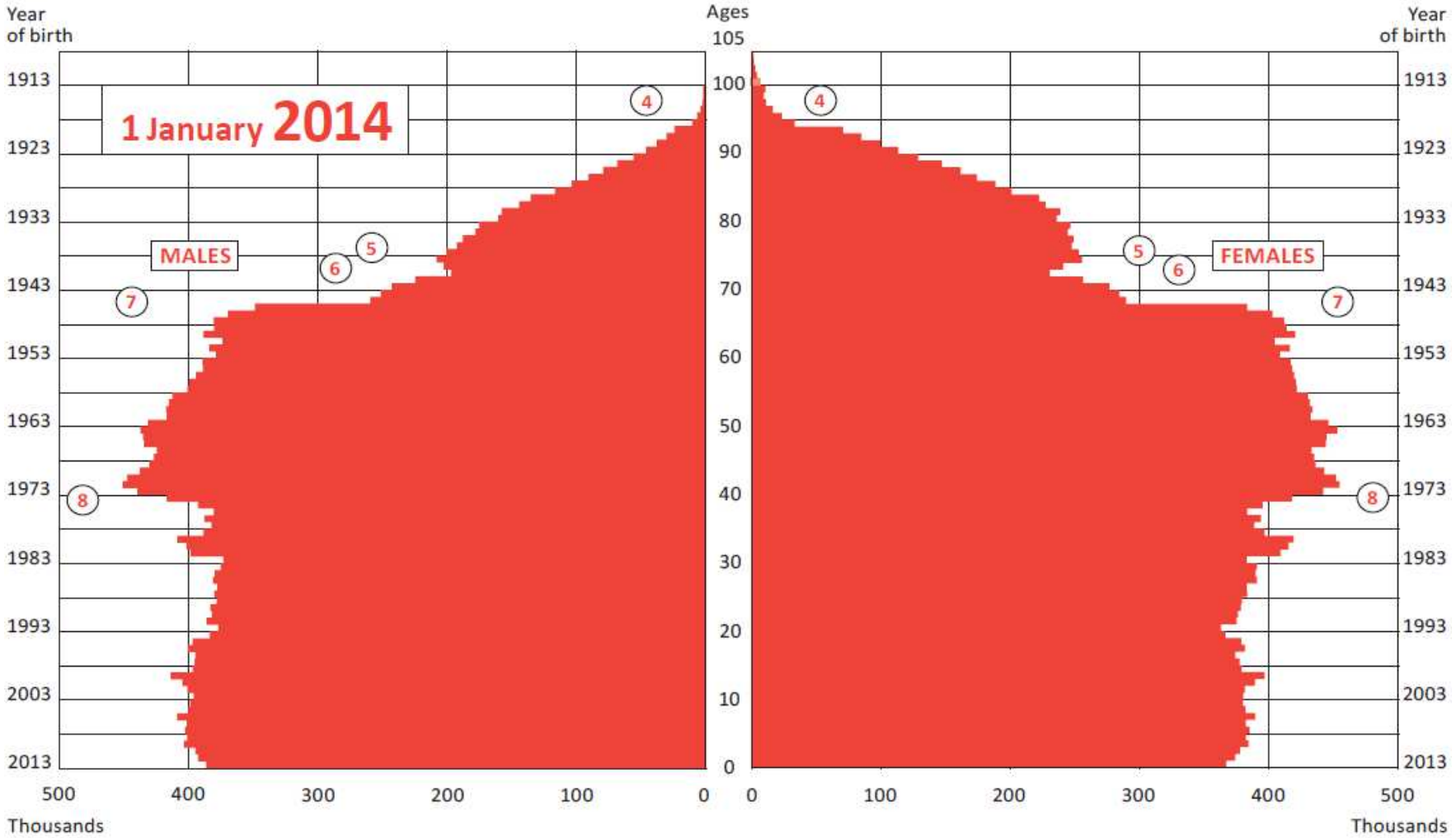
1. Shortfall of births due to the war of 1870
2. Exceptional infant mortality in 1911 due to a summer heat wave
3. Military losses of the 1914-1918 war
4. Shortfall of births due to the 1914-1918 war (depleted cohorts)
5. Depleted cohorts reach childbearing age
6. Shortfall of births due to the war of 1939-1945
7. Start of baby boom
8. End of baby boom

Population pyramid – France, 1974 (*Source*: Pison 2014)



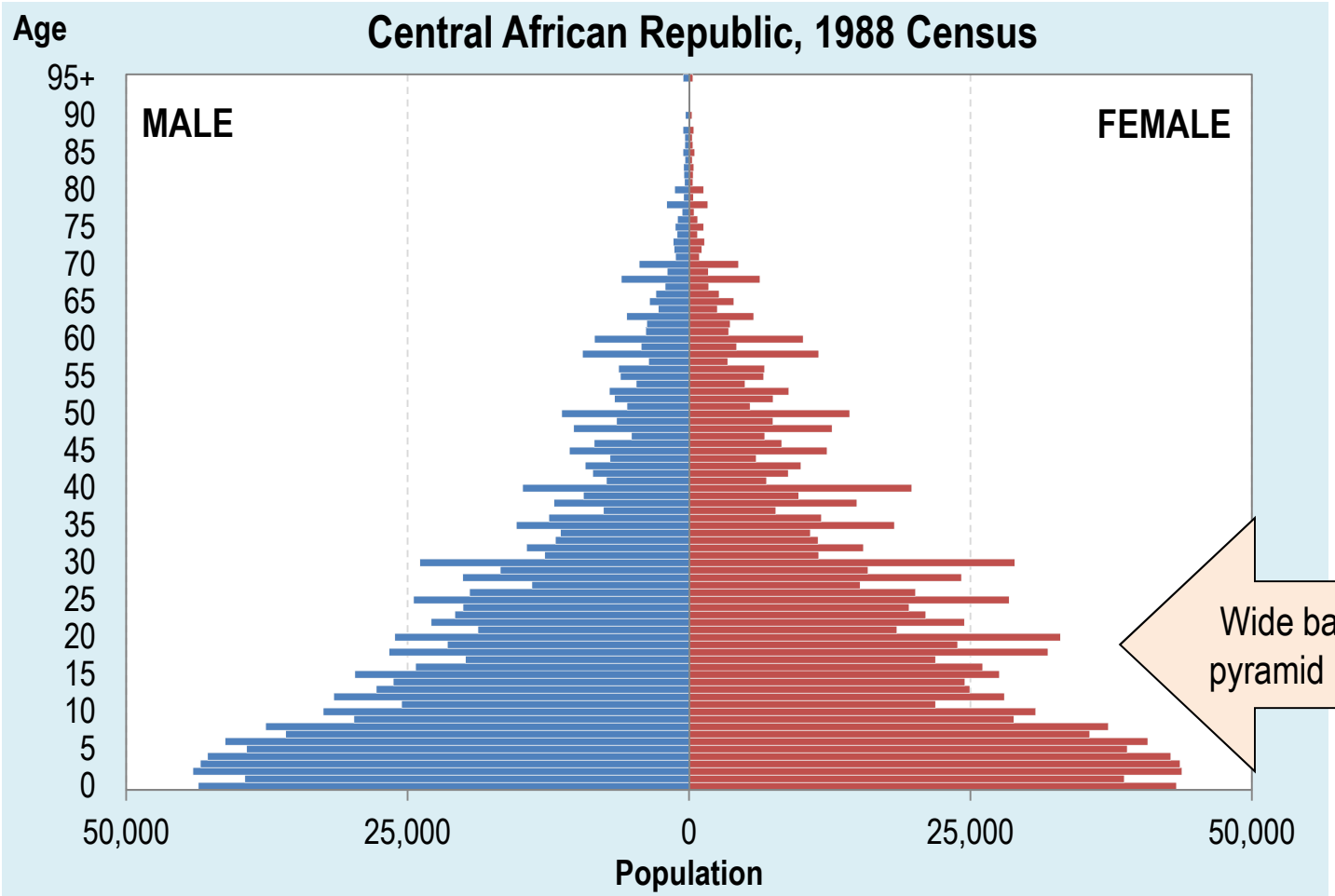
1. Shortfall of births due to the war of 1870
2. Exceptional infant mortality in 1911 due to a summer heat wave
3. Military losses of the 1914-1918 war
4. Shortfall of births due to the 1914-1918 war (depleted cohorts)
5. Depleted cohorts reach childbearing age
6. Shortfall of births due to the war of 1939-1945
7. Start of baby boom
8. End of baby boom

Population pyramid – France, 2014 (*Source*: Pison 2014)



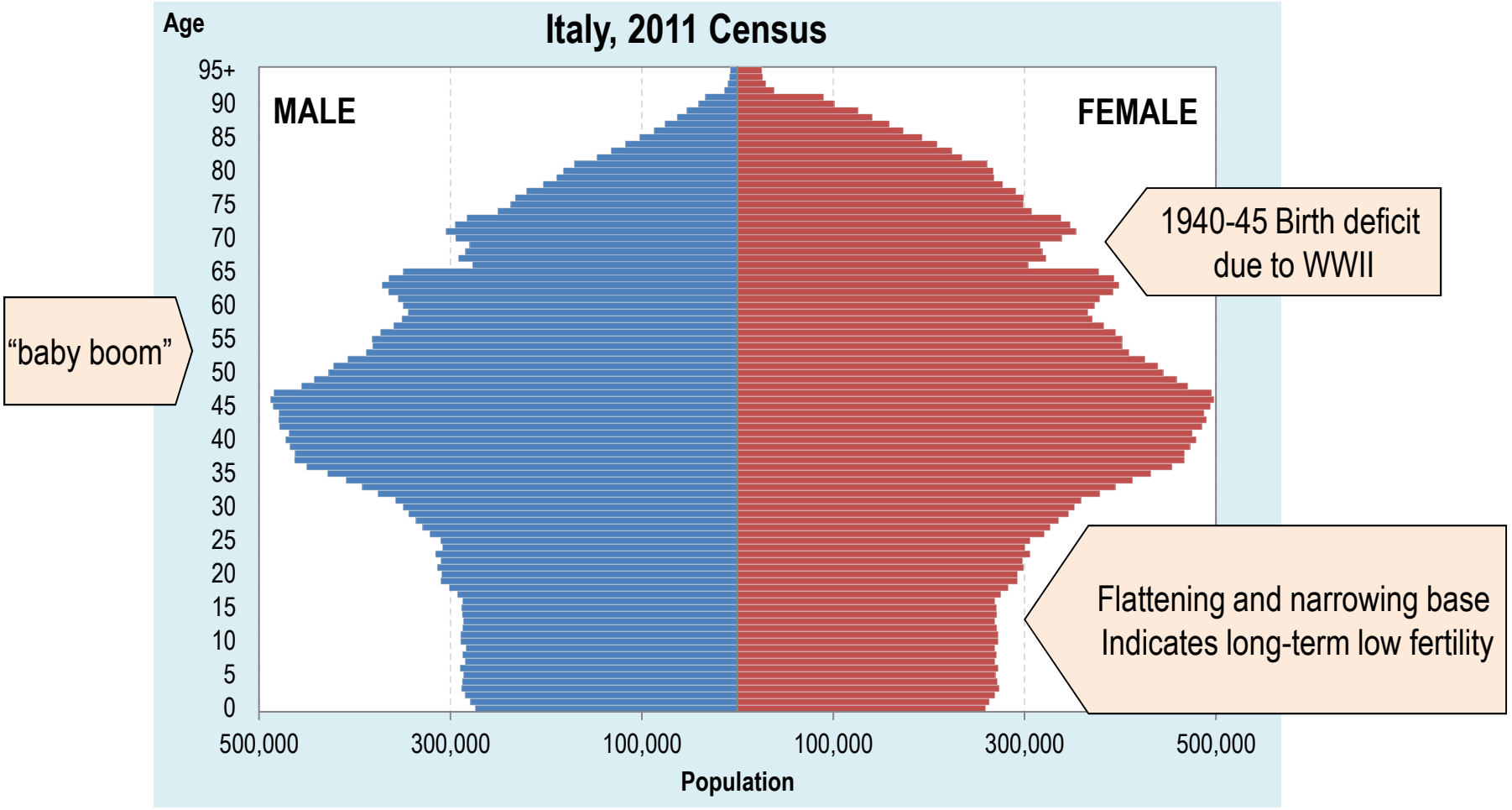
1. Shortfall of births due to the war of 1870
2. Exceptional infant mortality in 1911 due to a summer heat wave
3. Military losses of the 1914-1918 war
4. Shortfall of births due to the 1914-1918 war (depleted cohorts)
5. Depleted cohorts reach childbearing age
6. Shortfall of births due to the war of 1939-1945
7. Start of baby boom
8. End of baby boom

Population pyramid – High population growth

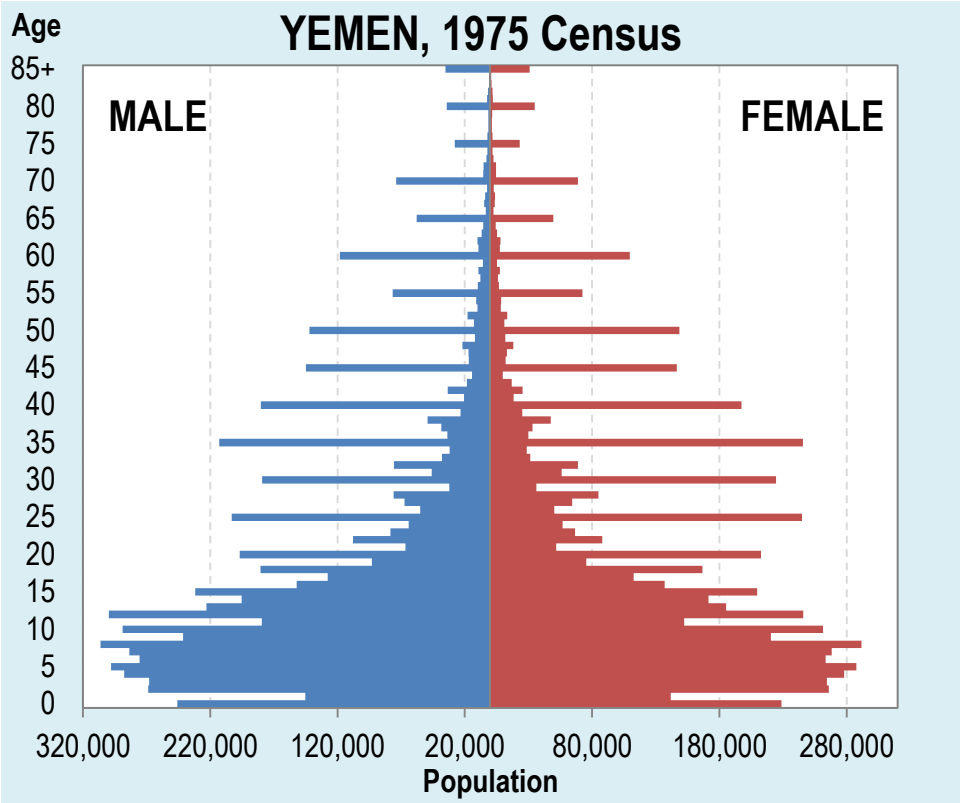


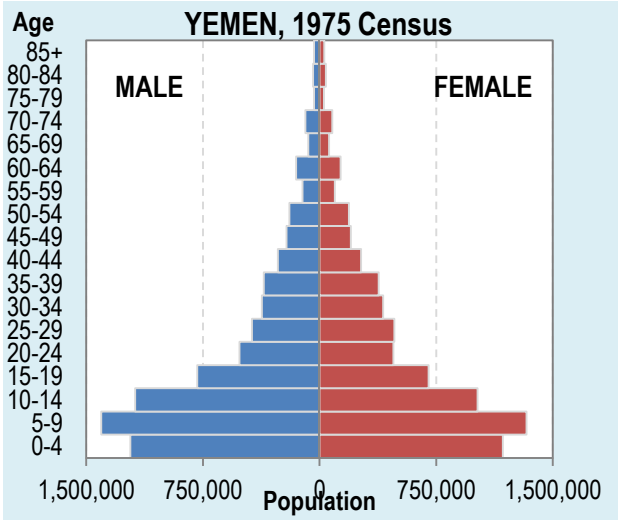
Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Population pyramid – Low population growth



Population pyramid – Detecting errors

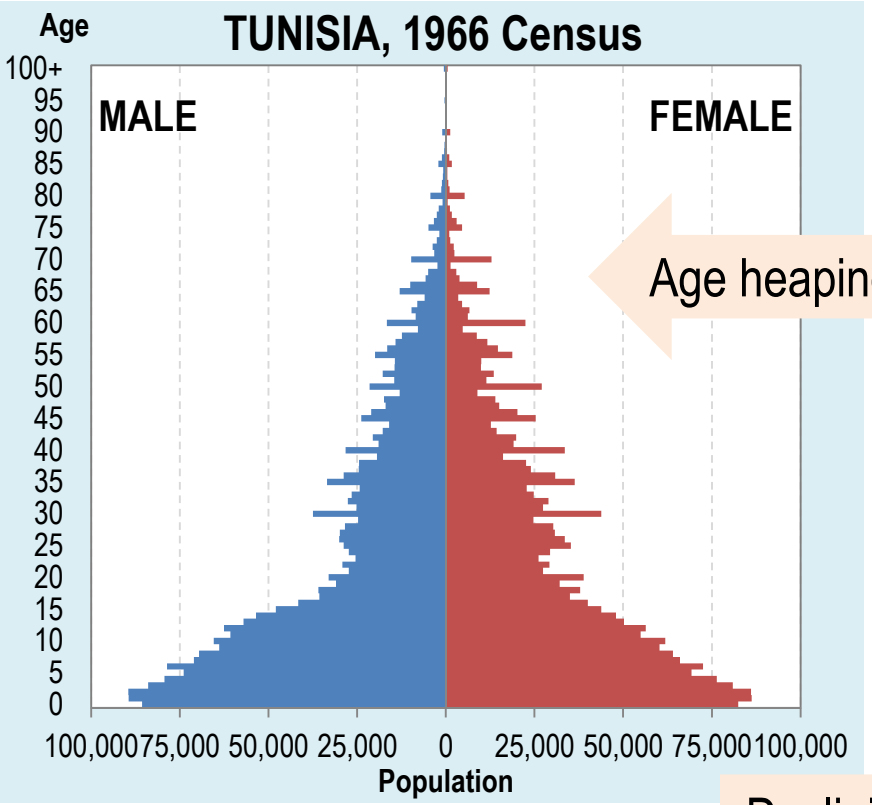




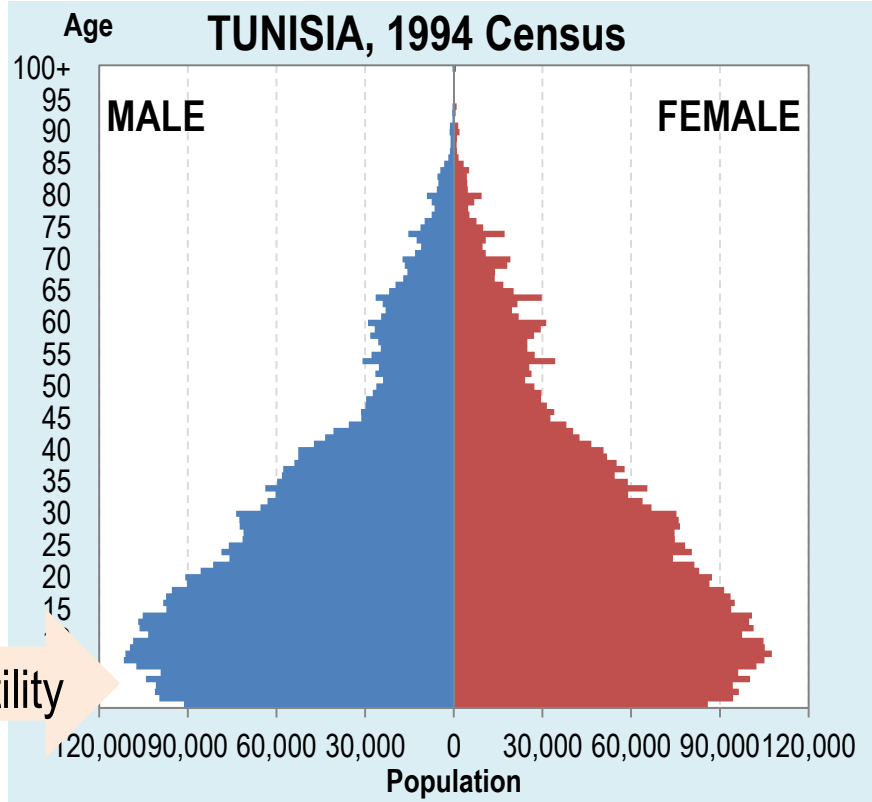
- Age misreporting errors (heaping) among adults
- Under enumeration of young children (< age 2)
- High fertility level
- Smaller population in 20-24 age group
  >> extraordinary events in 1950-55?
- Less men relative to women in 20-44 age group
  >> labor out-migration?

Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Population pyramid – Detecting errors



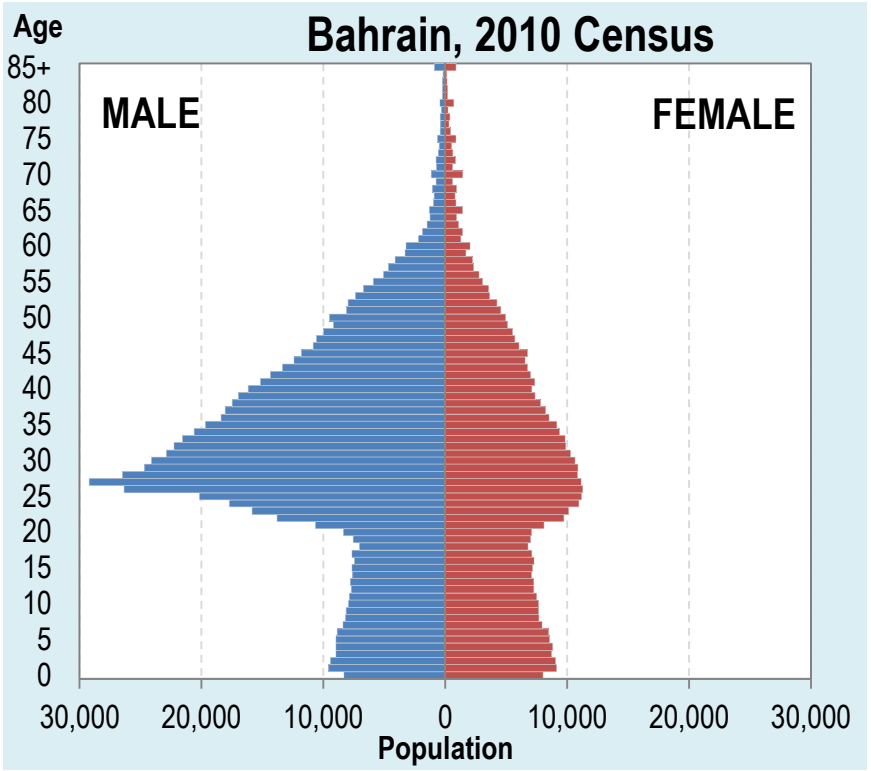
Age heaping

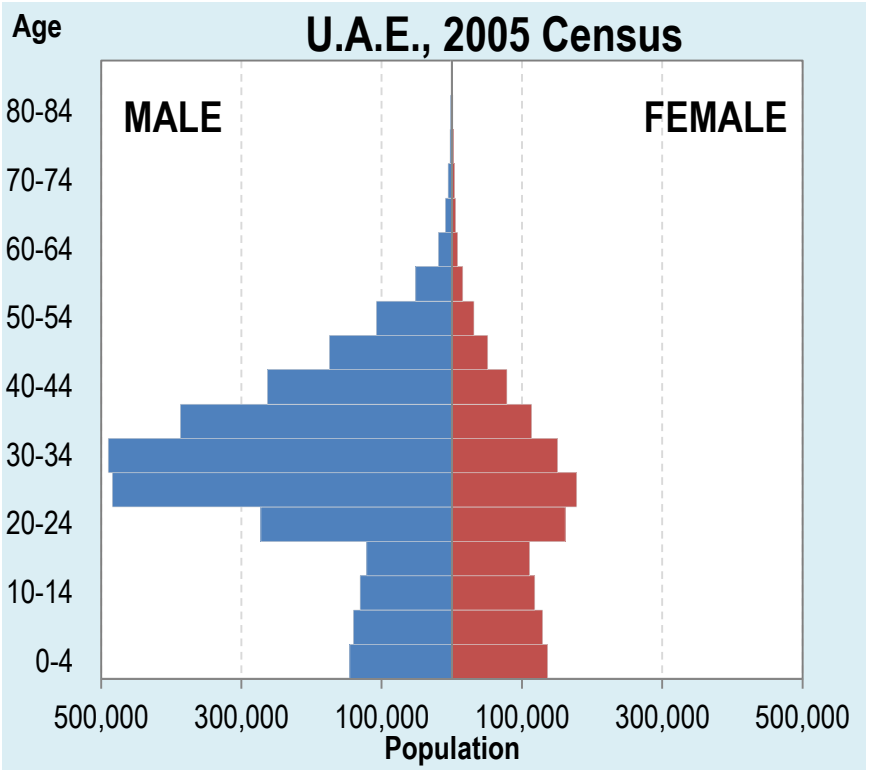


Declining fertility

Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Population pyramid – Detecting errors





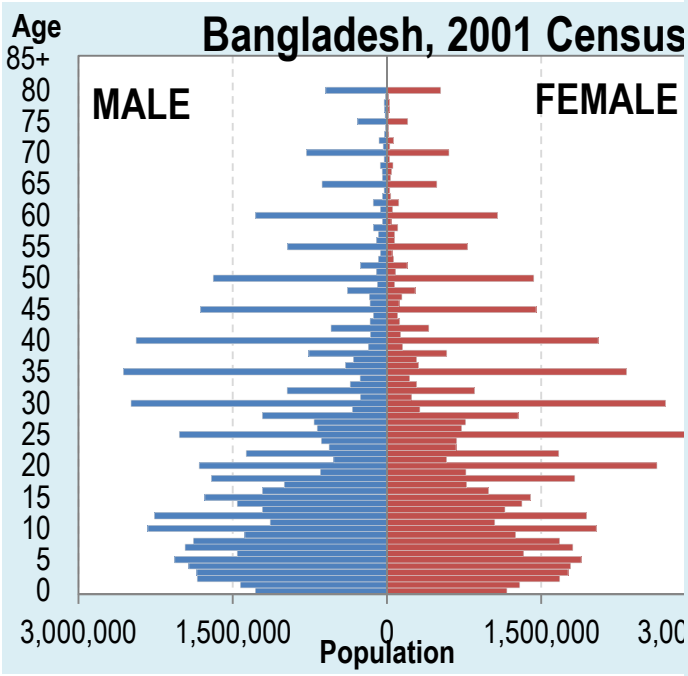
!!! Effect of Labor migration **!!!**

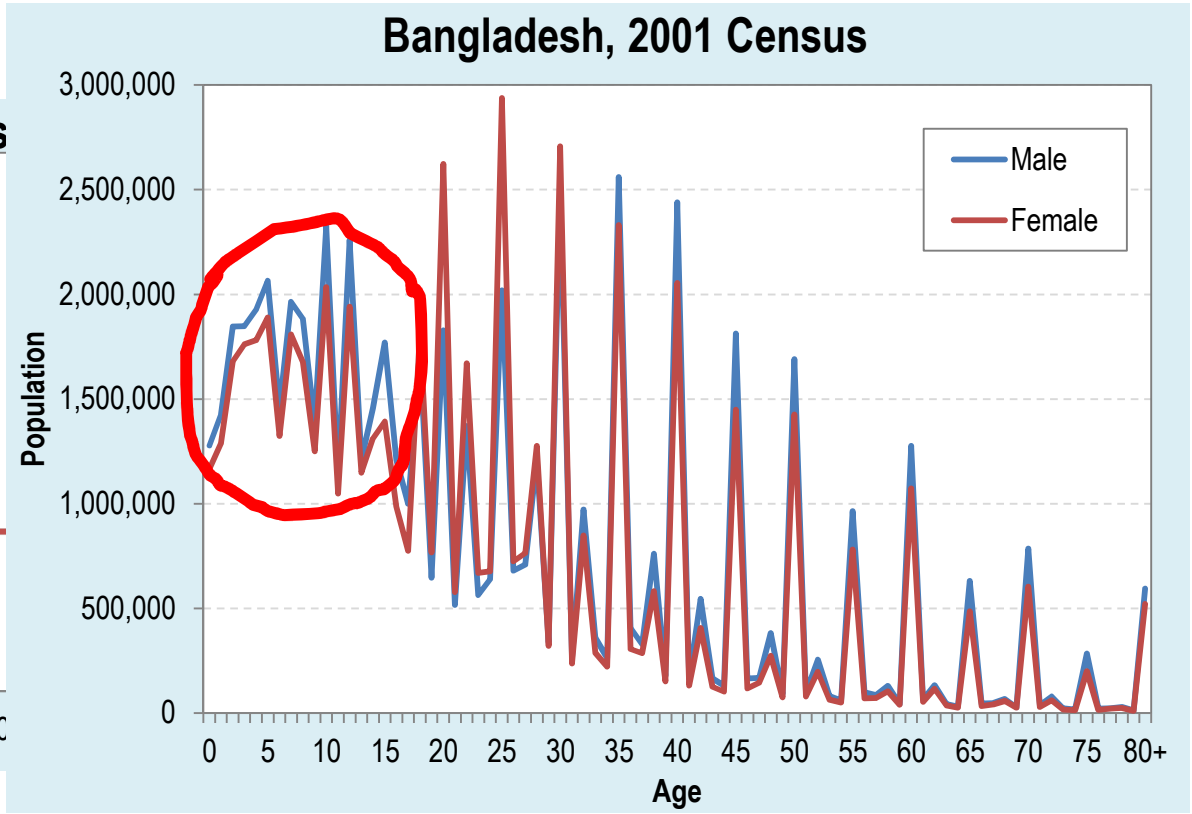
>> Importance of knowledge of the context

Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Population pyramid – Line instead of bars

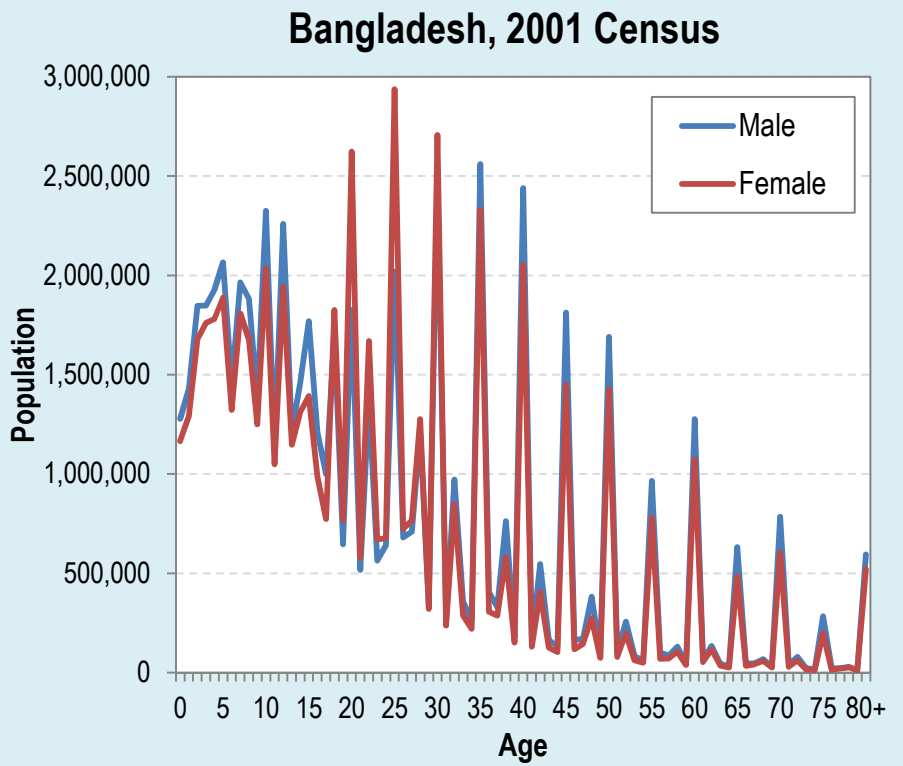
- Population Pyramid (bar chart)
  >> Not always easy to determine differences by sex
- Use of line chart

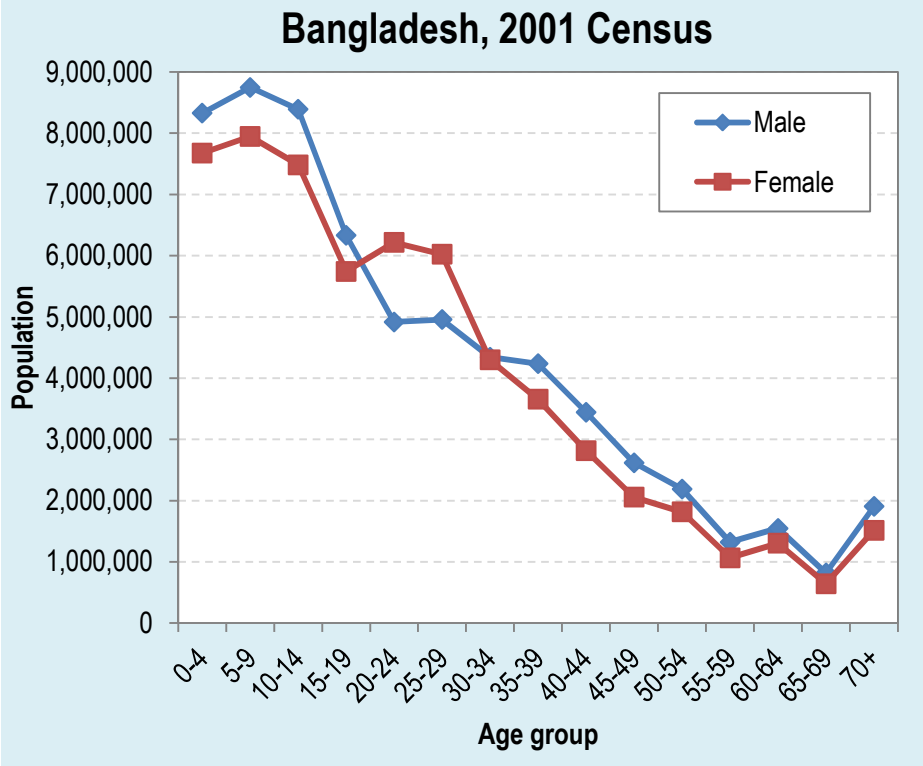




Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Population pyramid – Line instead of bars





Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Basic graphical methods – Graphical cohort analysis

- Tracking actual cohorts over multiple censuses
- The size of each cohort should decline over each census due to mortality (with no significant international migration)
- The age structure (the lines) for censuses should follow the same pattern in the absence of census errors
- An important advantage **>>** possible to evaluate the effects of extraordinary events and other distorting factors by following actual cohorts over time

Graphical cohort analysis – Example (1)

Mozambique, 1997 and 2007 Censuses

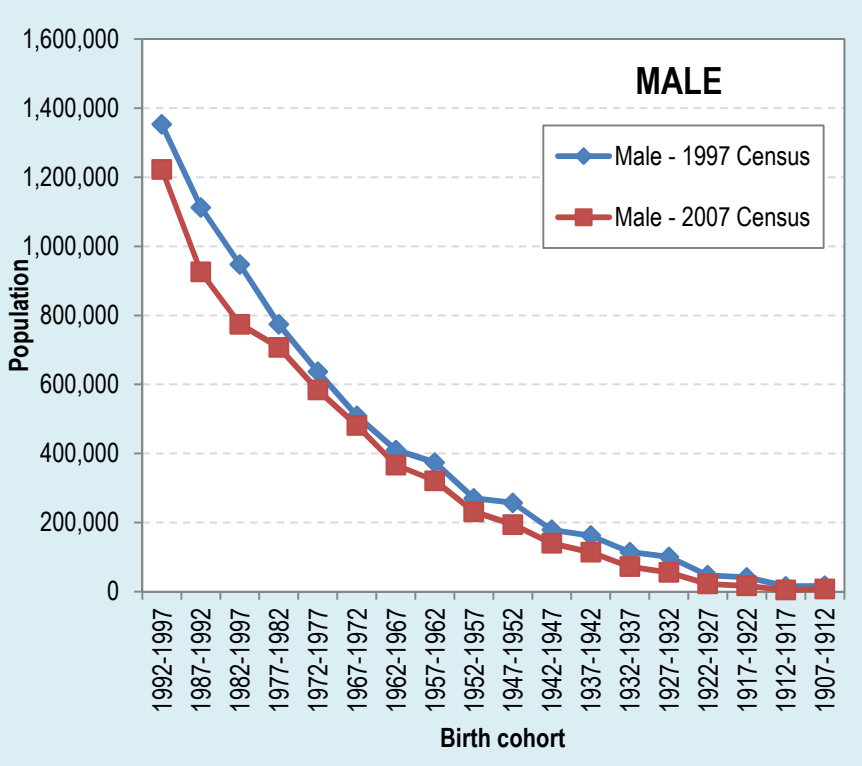
	1997 Census		2007 Census		Birth Cohort
	Male	Female	Male	Female	
0-4	1,353,206	1,388,350			2002-2007
5-9	1,112,321	1,113,675			1997-2001
10-14	947,236	878,429	1,222,668	1,183,939	1992-1997
15-19	774,327	854,078	925,729	991,323	1987-1992
20-24	637,113	827,614	774,413	986,526	1982-1997
25-29	509,109	654,465	707,603	841,416	1977-1982
30-34	410,148	477,562	583,689	667,865	1972-1977
35-39	373,813	428,395	481,396	556,191	1967-1972
40-44	270,046	303,147	366,518	389,087	1962-1967
45-49	257,070	282,098	321,236	328,660	1957-1962
50-54	178,902	212,060	231,232	283,288	1952-1957
55-59	162,122	174,234	194,011	208,657	1947-1952
60-64	114,335	125,096	140,146	159,557	1942-1947
65-69	100,425	109,288	113,840	127,794	1937-1942
70-74	47,407	50,607	72,288	81,329	1932-1937
75-79	41,529	42,858	55,448	61,012	1927-1932
80-84	15,305	17,326	22,417	28,278	1922-1927
85-89			16,576	19,448	1917-1922
90-94			4,803	5,883	1912-1917

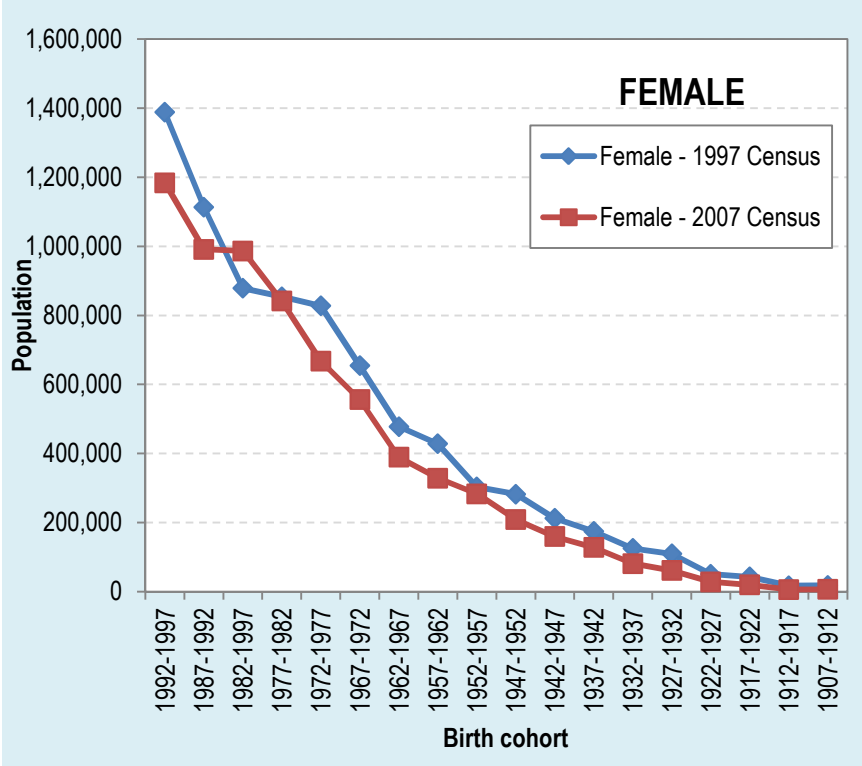
- Data is organized by birth cohort
- Exclude open-ended age category
- People who were born in the same years are compared in the analysis

Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Graphical cohort analysis – Example (1)

Mozambique, 1997 and 2007 Censuses

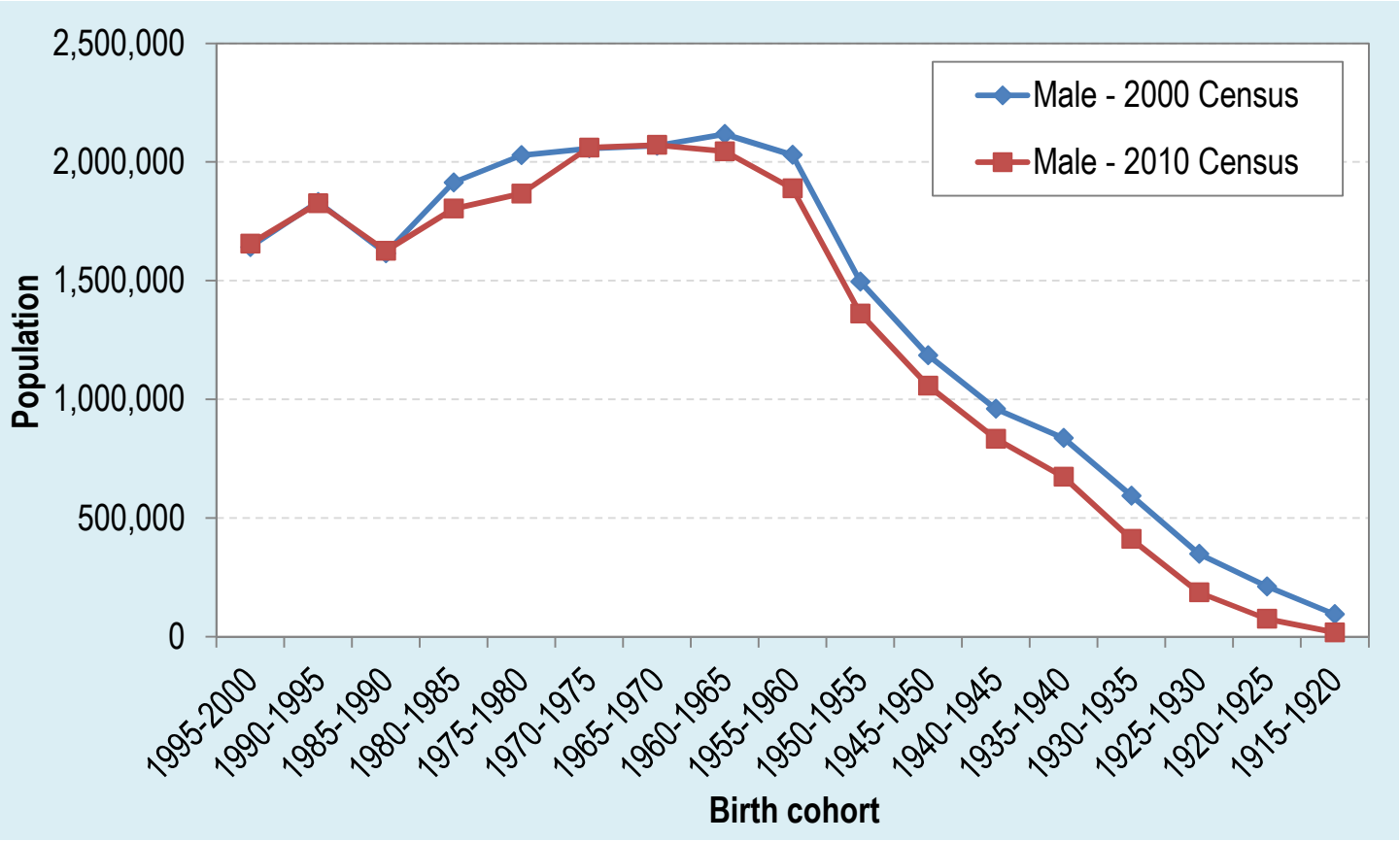




Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Graphical cohort analysis – Example (2)

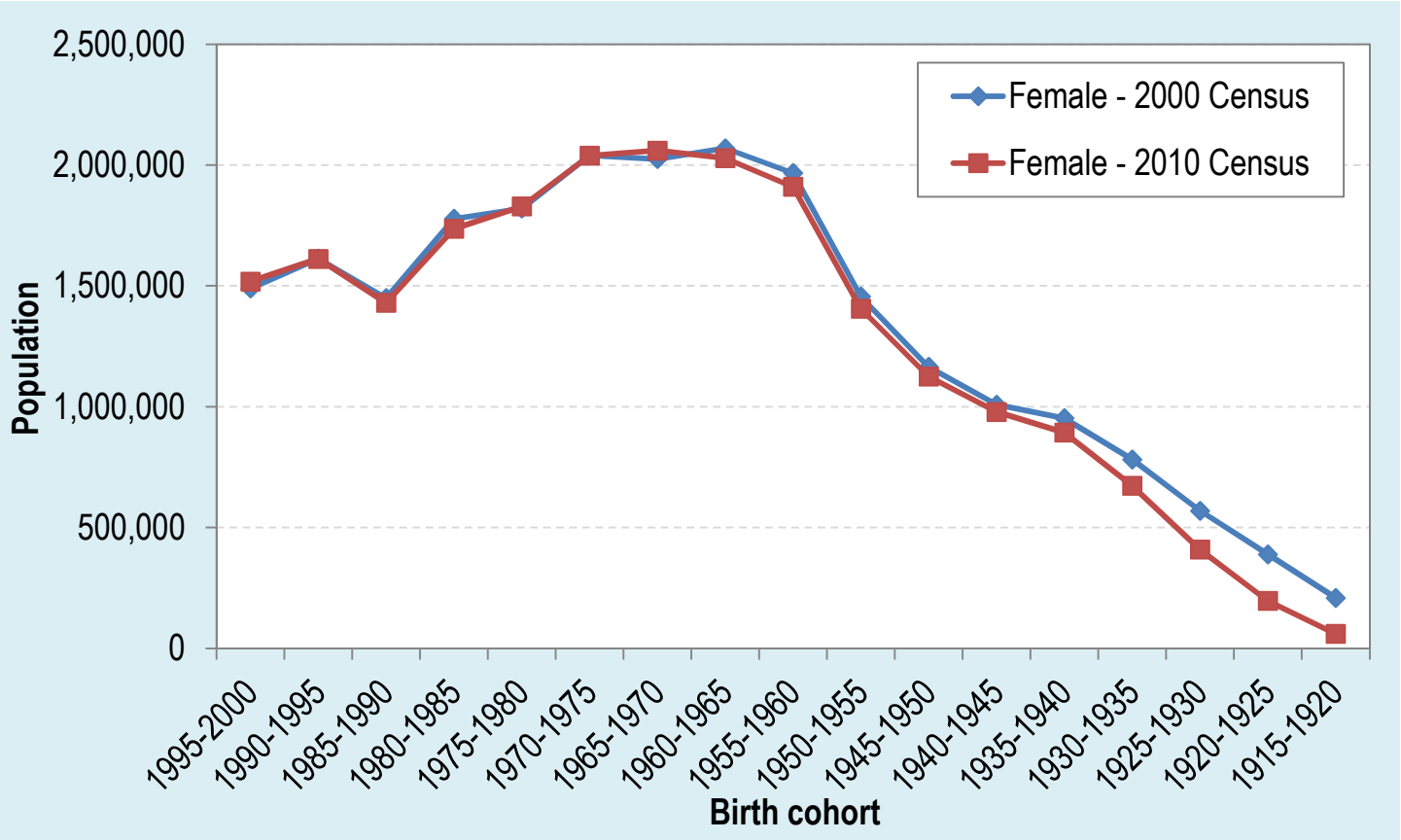
Republic of Korea, 2000 and 2010 Censuses, Men



Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Graphical cohort analysis – Example (2)

Republic of Korea, 2000 and 2010 Censuses, Women



Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Age ratios (1)

- In the absence of sharp changes in fertility or mortality, significant levels of migration or other distorting factors, the enumerated size of a particular cohort should be approximately equal to the average size of the immediately preceding and following cohorts
- The age ratio for a particular cohort to the average of the counts for the adjacent cohorts should be approximately equal to 1 (or 100 if multiplied by a constant of 100)
- Significant departures from this “expected” ratio indicate either the presence of census error in the census enumeration or of other factors

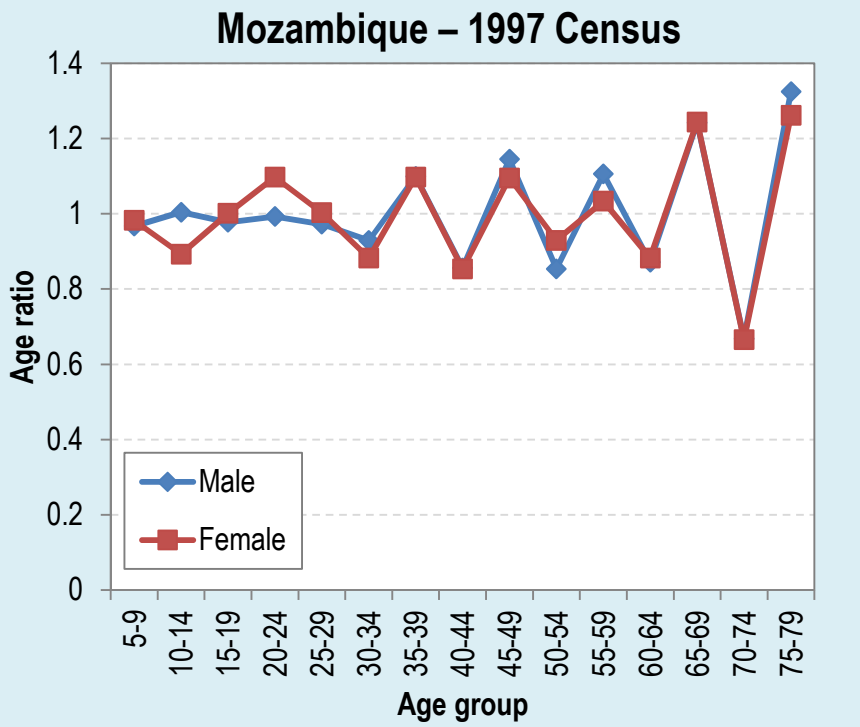
Age ratios (2)

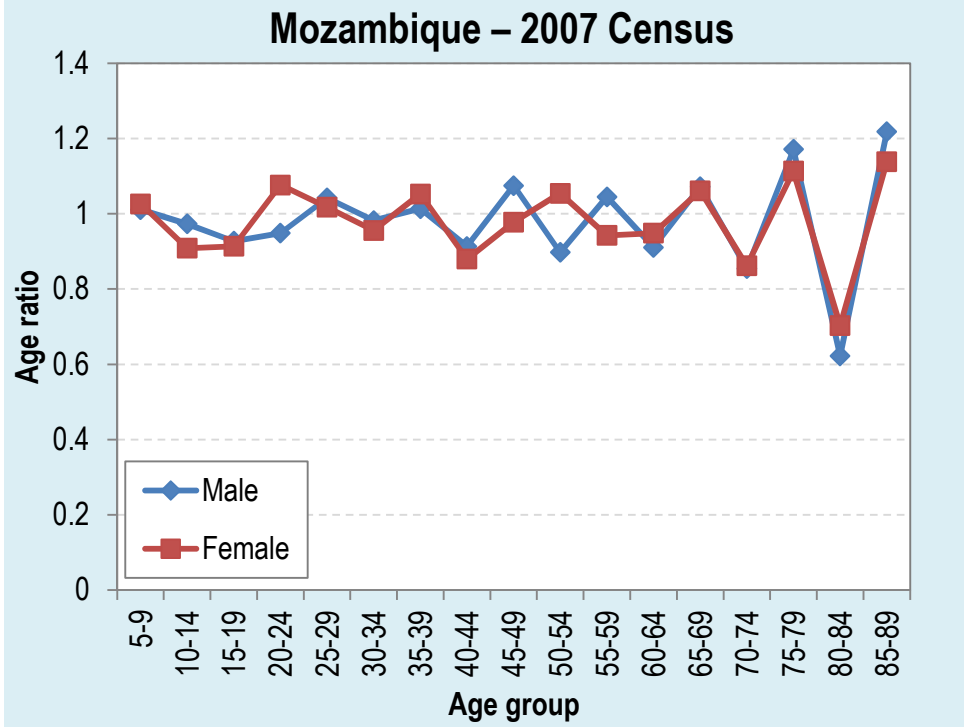
Age ratio for the age category x to x+4

$$ {}_5AR_x = \frac{2 * {}_5P_x}{{}_5P_{x-n} + {}_5P_{x+n}} $$

${}_5AR_x$ = The age ratio for the age group x to $x + 4$
${}_5P_x$ = The enumerated population in the age category x to $x + 4$
${}_5P_{x-n}$ = The enumerated population in the adjacent lower age category
${}_5P_{x+n}$ = The enumerated population in the adjacent higher age category

Age ratios (3) – example





Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Sex ratios (1) - calculation

Sex ratio by age group

$$\text{Sex Ratio} = \frac{{}_5P_x^m}{{}_5P_x^f} \qquad \textit{or} \qquad \text{Sex Ratio} = \frac{{}_5P_x^m}{{}_5P_x^f} \cdot 100$$

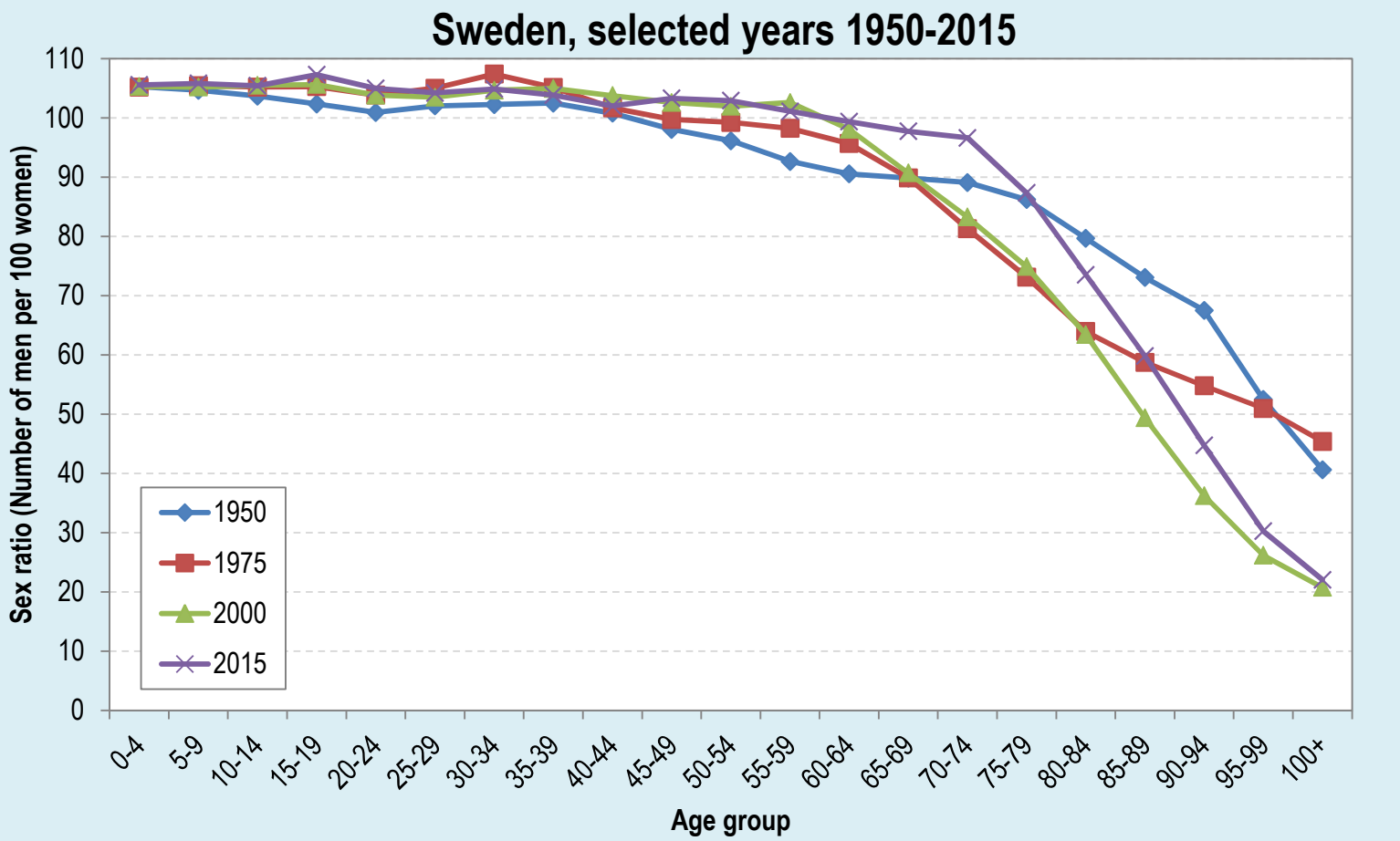
Where

${}_5P_x^m$ = Male population enumerated in a specific age group

${}_5P_x^f$ = Female population enumerated in the same age group

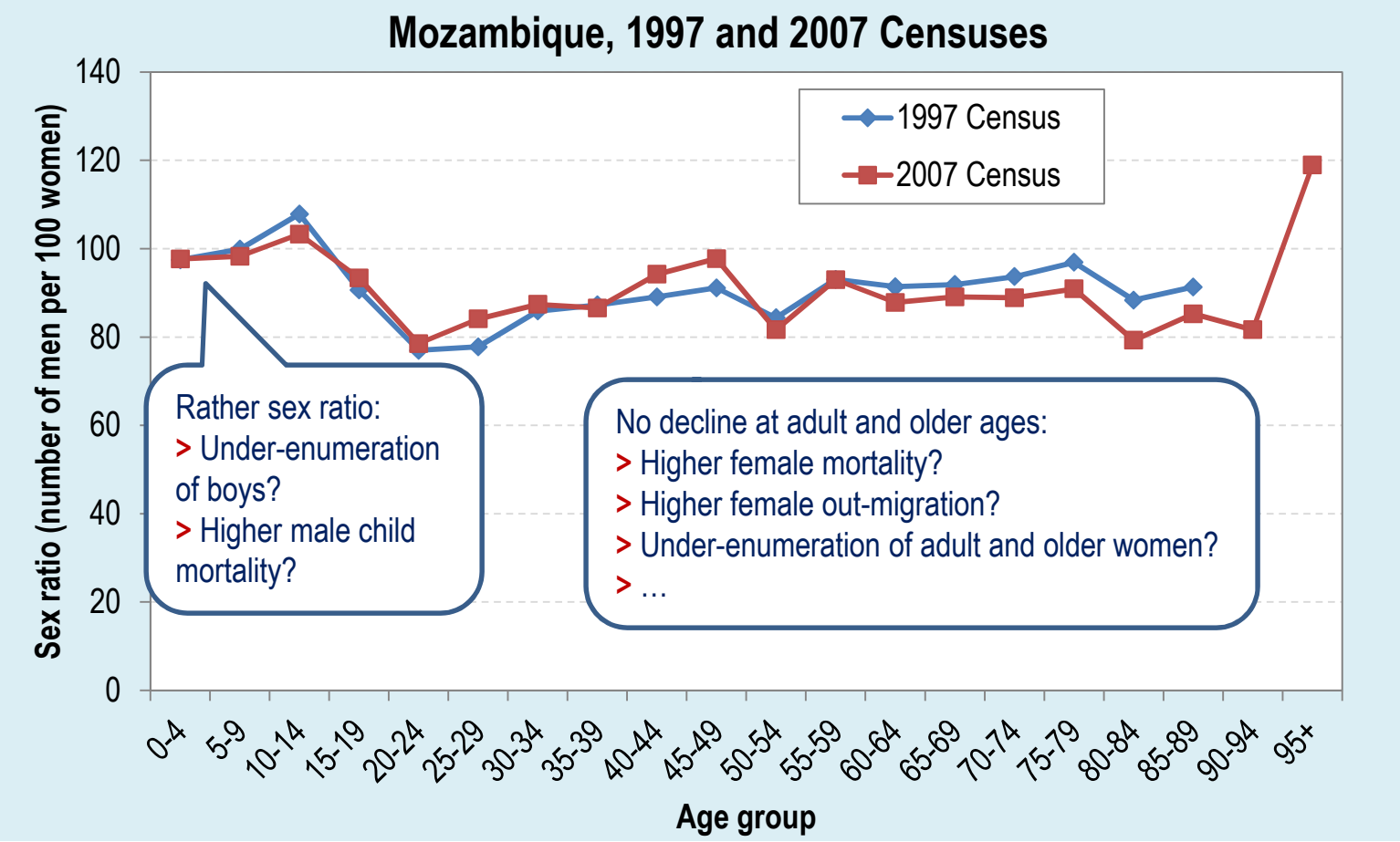
Value of sex ratio	Interpretation
1	Same number of men and women in a given age group
Above 1	More men than women in a given age group
Below 1	Less men than women in a given age group

Sex ratios – Example (Expected sex ratio by age)



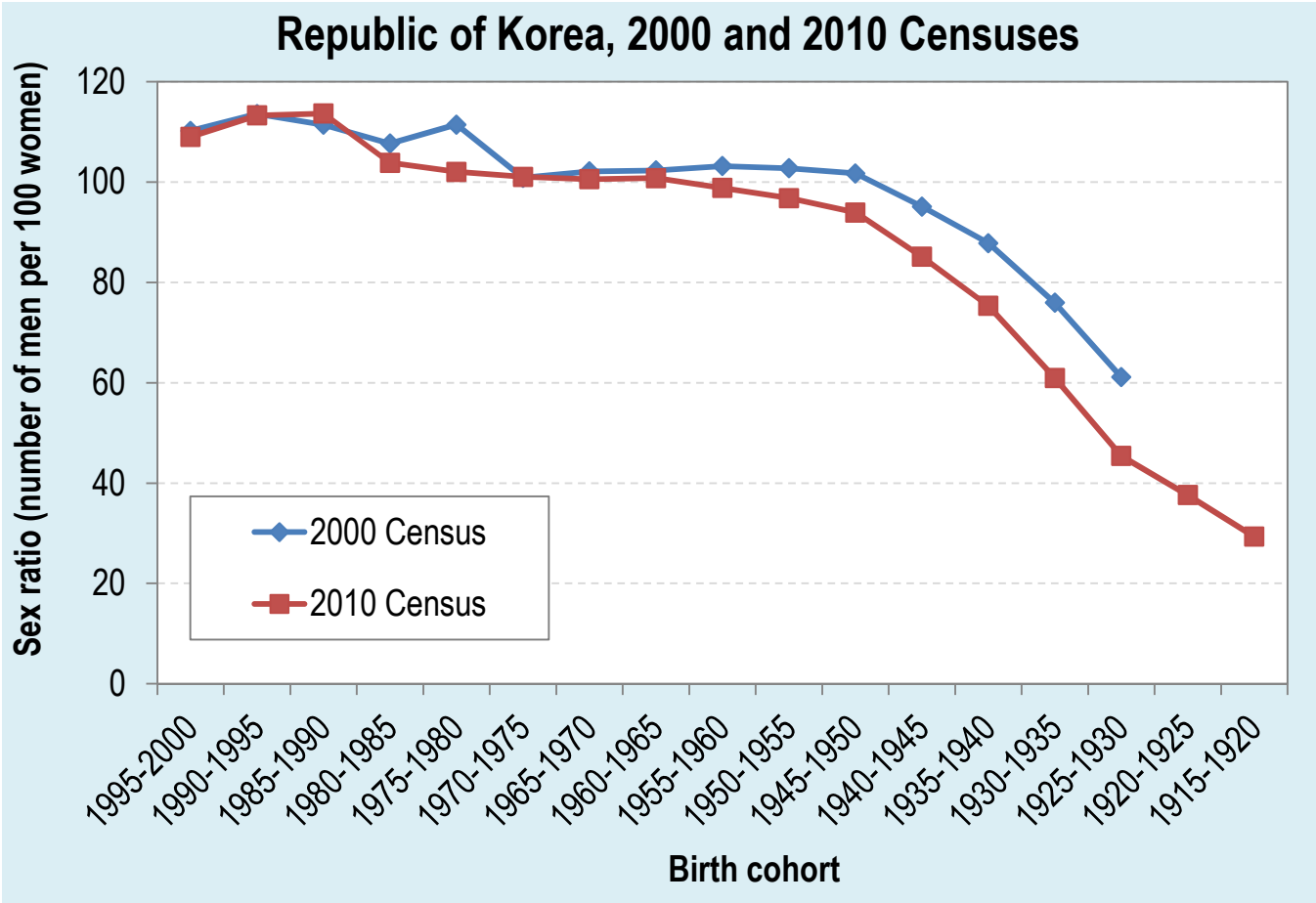
Source: United Nations Population Division, *The World Population Prospects: The 2015 Revision*

Sex ratios – Example



Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Sex ratios – Cohort analysis

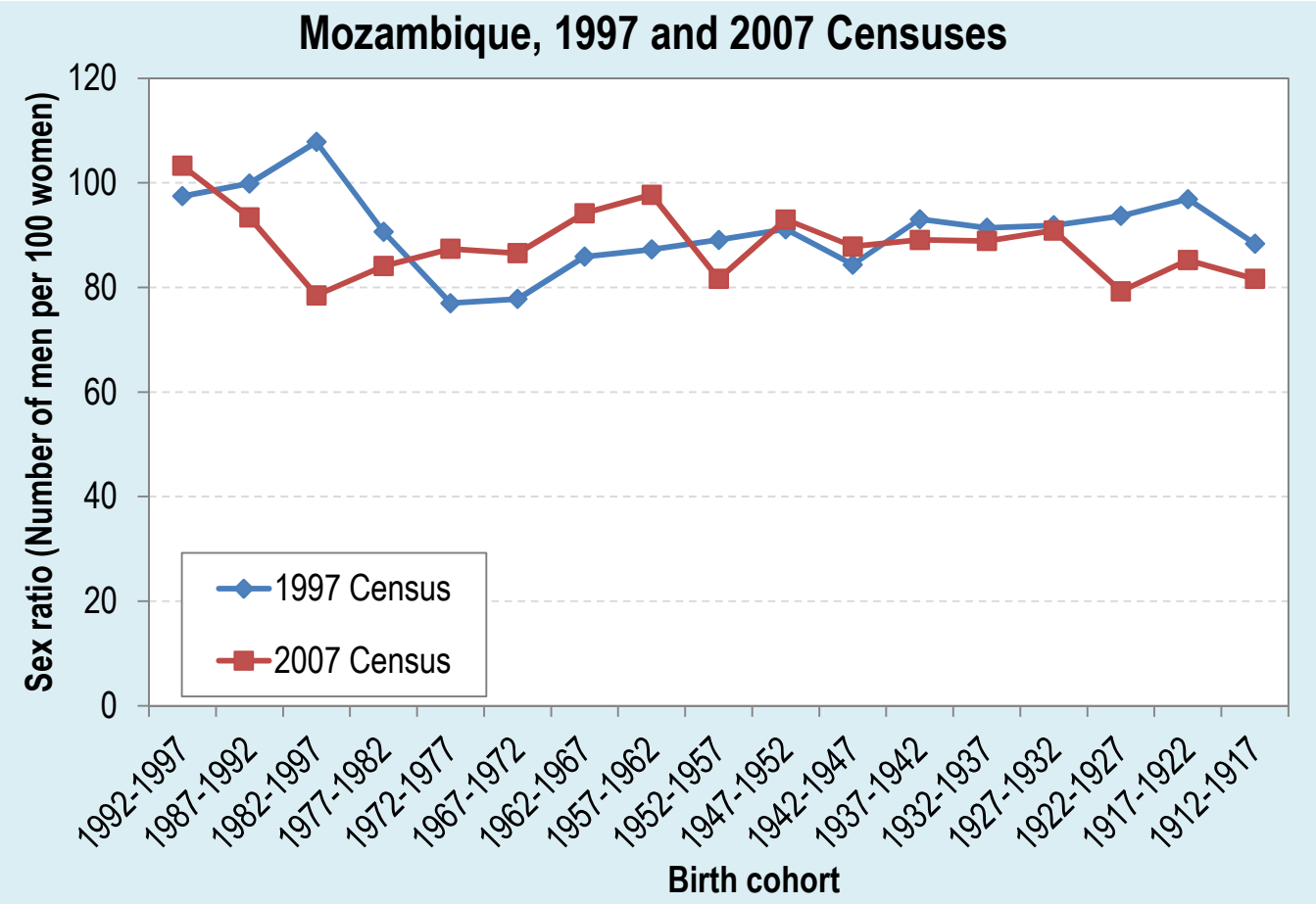


> Excess of men in 1975-1980 birth cohort in 2000 census only

> Other age groups: as expected

Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Sex ratios – Cohort analysis



> Very different patterns found in each census

> Some consistencies for birth cohorts 1947-1952 to 1927-1932

Source: United Nations Statistics Division, *Demographic Yearbook Statistical Database*

Summary indices – Whipple's Index

- Developed to reflect preference for or avoidance of a particular terminal digit (or of each terminal digit)
- If heaping on terminal digits "0" and "5" is measured:

$$W = \frac{(P_{25}+P_{30}+P_{35}+ \ldots + P_{55}+P_{60})}{\frac{1}{5} \cdot (P_{23}+P_{24}+P_{25}+ \ldots + P_{61}+P_{62})} \cdot 100$$

- W ranges between 100, representing no preference for "0" or "5" and 500, indicating that only digits "0" and "5" were reported
- W is usually computed on the age range 23 to 62

  >> childhood and older ages are often excluded because they are more strongly affected by other types of errors of reporting than by preference for specific terminal digits

Source: Shryock and Siegel (1980)

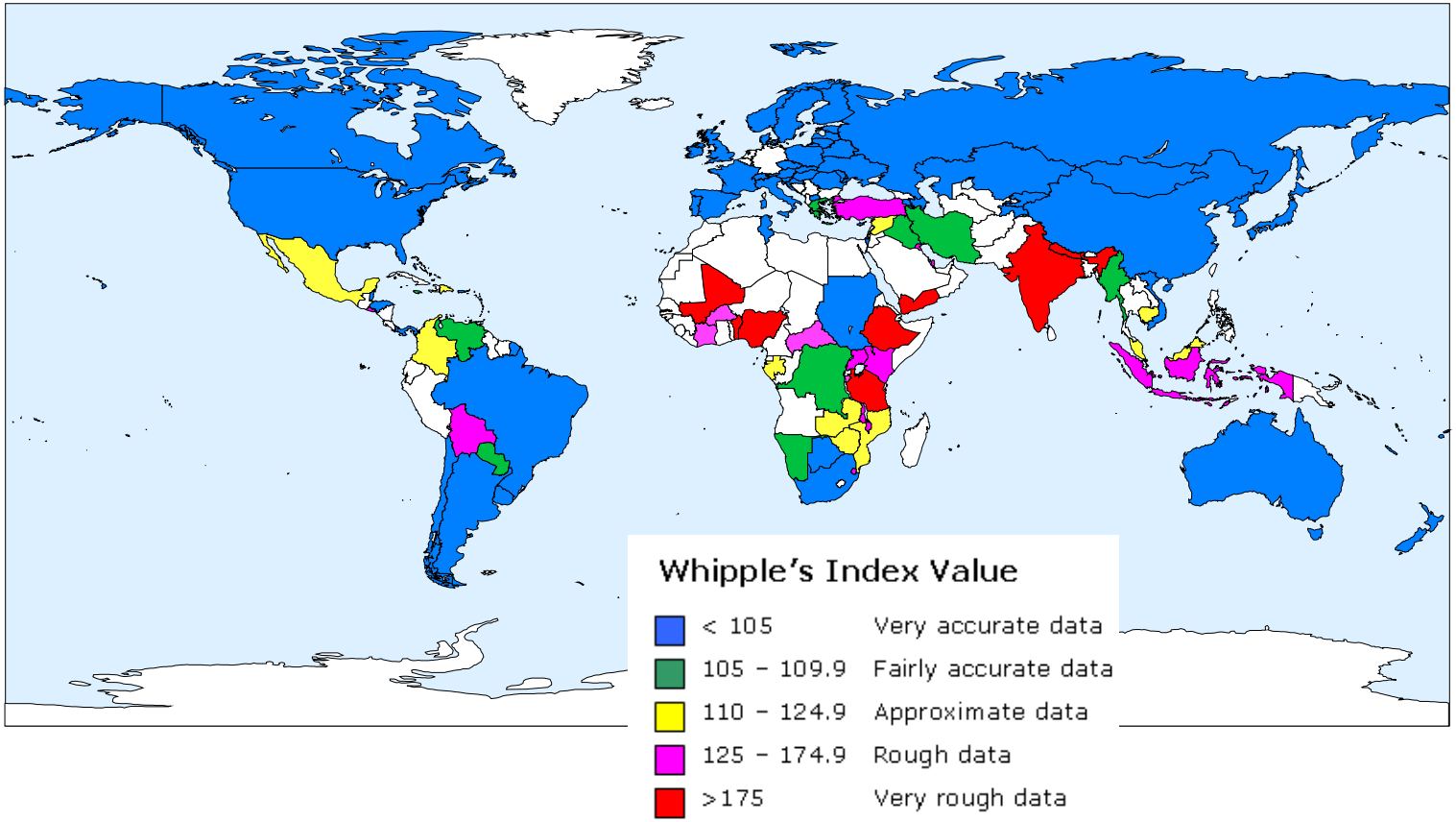
Whipple's Index (3)

The index can be summarized through the following categories:

Value of Whipple's index (W)	Interpretation
<= 105	Highly accurate data
105 – 109.9	Fairly accurate data
110 – 124.9	Approximate data
125 – 174.9	Rough data
>= 175	Very rough data

>> these values must be interpreted keeping in mind the past and present context

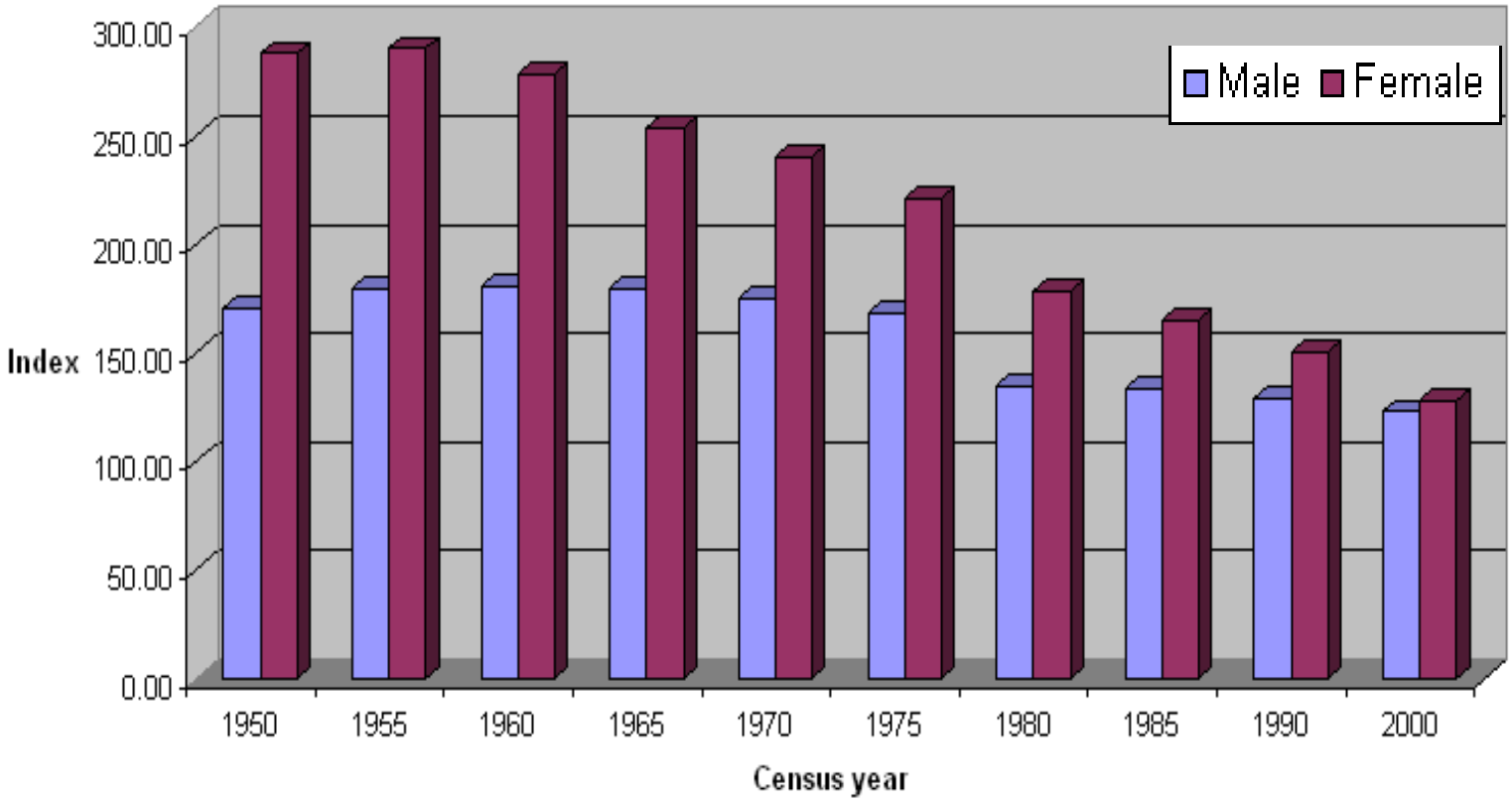
Whipple's index around the world (most recent census, 1985-2003)



Source: UN *Demographic Yearbook* special issue on age heaping: http://unstats.un.org/unsd/demographic/products/dyb/dybcens.htm

Improvement in the accuracy of age reporting over time

Whipple's Index, Turkey 1950-2000



Summary indices – Myers` Blended Index

- It is conceptually similar to Whipple's index, except that the index considers preference (or avoidance) of age ending in each of the digits 0 to 9 in deriving overall age accuracy score
- The theoretical range of Myers' Index is from 0 to 90, where 0 indicates no age heaping and 90 indicates the extreme case where all recorded ages end in the same digit

Conclusion: Uses and limitations

- Assessment of the age and sex structure of a population enumerated is typically the first step taken in evaluating a data collection operation by means of demographic methods
- Demographic methods provide:
  - A quick and inexpensive indication of the general quality of data
  - Evidence on the specific segments of the population in which the presence of error is likely
  - “Historical” information which may be useful for interpreting the results of evaluation studies based on other methods, and in determining how the collected data should be adjusted for use in demographic analyses

Conclusion: Uses and limitations

- The major limitation of age and sex structure analysis is that it is not possible to derive separate numerical estimates of the magnitude of coverage and content error on the basis of such analyses alone
- It is often possible to assess particular types of errors which are likely to have affected the population counts for particular segments of the population. Estimates of coverage error from other sources often are required to verify these observations

References

Pison, G. (2014), "1914-2014 : A century of change in the French population pyramid", *Population & Societies* No. 509 (March 2014)

Shryock, H. S. and Siegel, J. S. (1980), *The Methods and Materials of Demography*, Washington, D.C.: U.S. Government Printing Office.

Moultrie T.A., R.E. Dorrington, A.G. Hill, K. Hill, I.M. Timæus and B. Zaba (eds), (2013), *Tools for Demographic Estimation,* Paris: International Union for the Scientific Study of Population, available online at: http://demographicestimation.iussp.org/

U.S. Census Bureau (n.d.), *Population Analysis System (PAS),* Washington, D.C., U.S. Census Bureau, available online at: http://www.census.gov/population/international/software/uscbtoolsdownload.html