Evaluation of Age and Sex Distribution Data

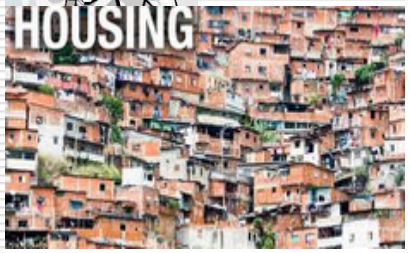
United Nations Statistics Division

Evaluation method of age and sex distribution data

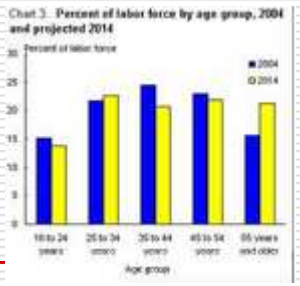
- Basic tools
  - Graphical analysis
    - Population pyramids
    - Graphical cohort analysis
  - Age and sex ratios
  - Summary indices of error in age-sex data
    - Whipple's index
    - Myers' Blended Method
- Uses of consecutive censuses

Focus of the presentation

Importance of age-sex structures



- Planning purposes – health services, sales programs, school, voting, 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.

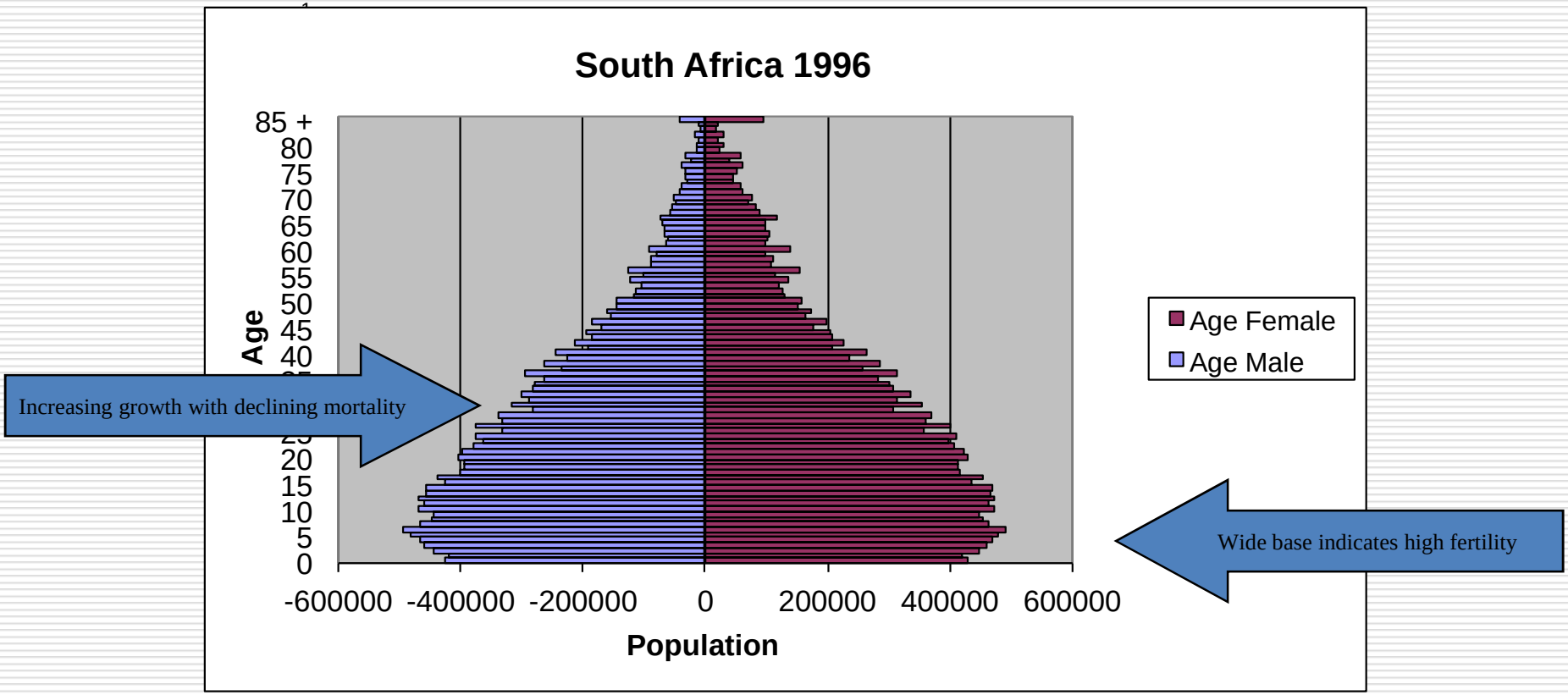
Approaches to collecting age and its impact on quality

- 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

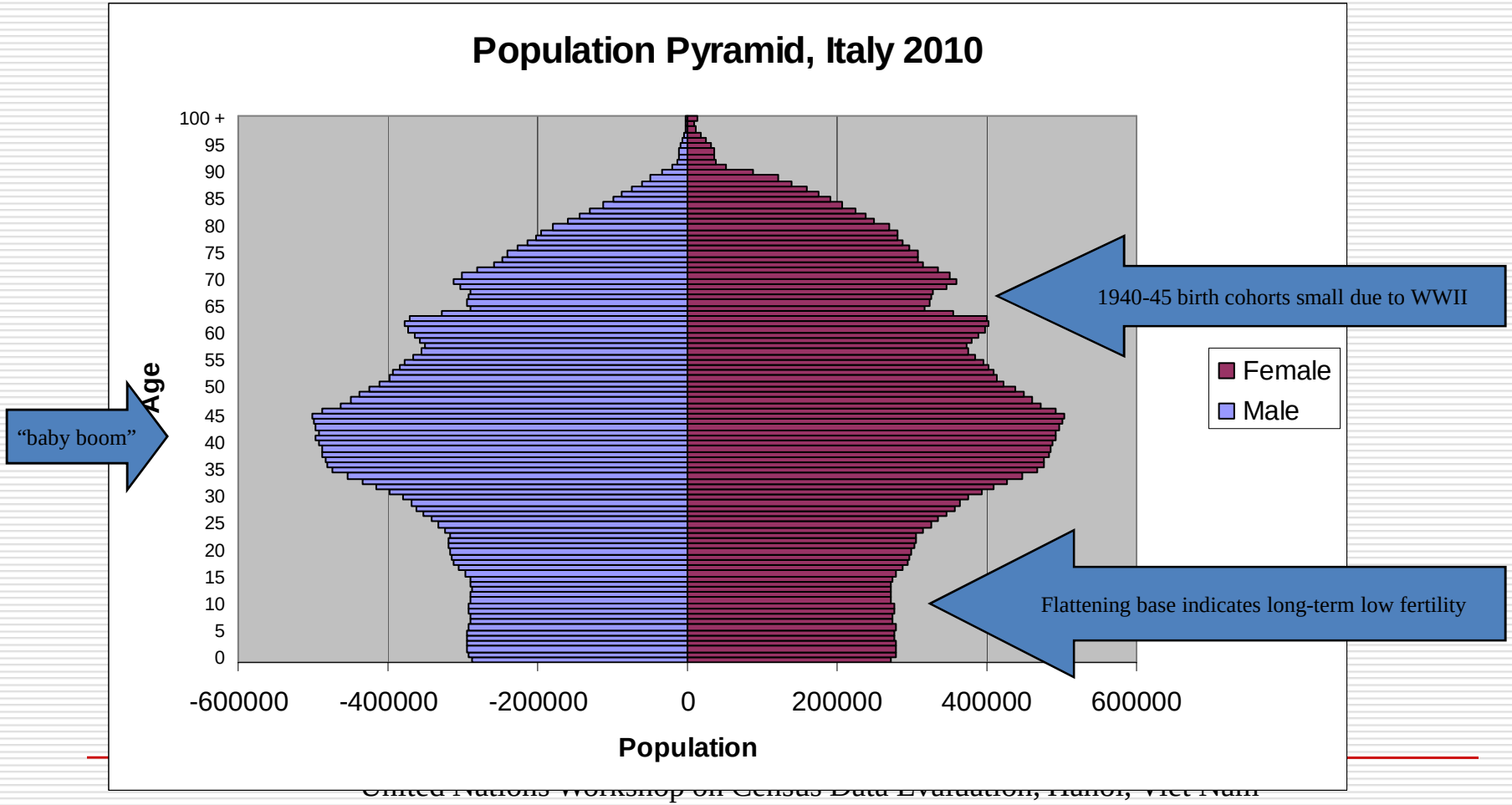
- Basic procedure for assessing the quality of census data on age and sex
- Displays the size of population enumerated in each age group (or cohort) by sex
- The base of the pyramid is mainly determined by the level of fertility in the population, while how fast it converges to peak is determined by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

Population pyramid (1) – high population growth

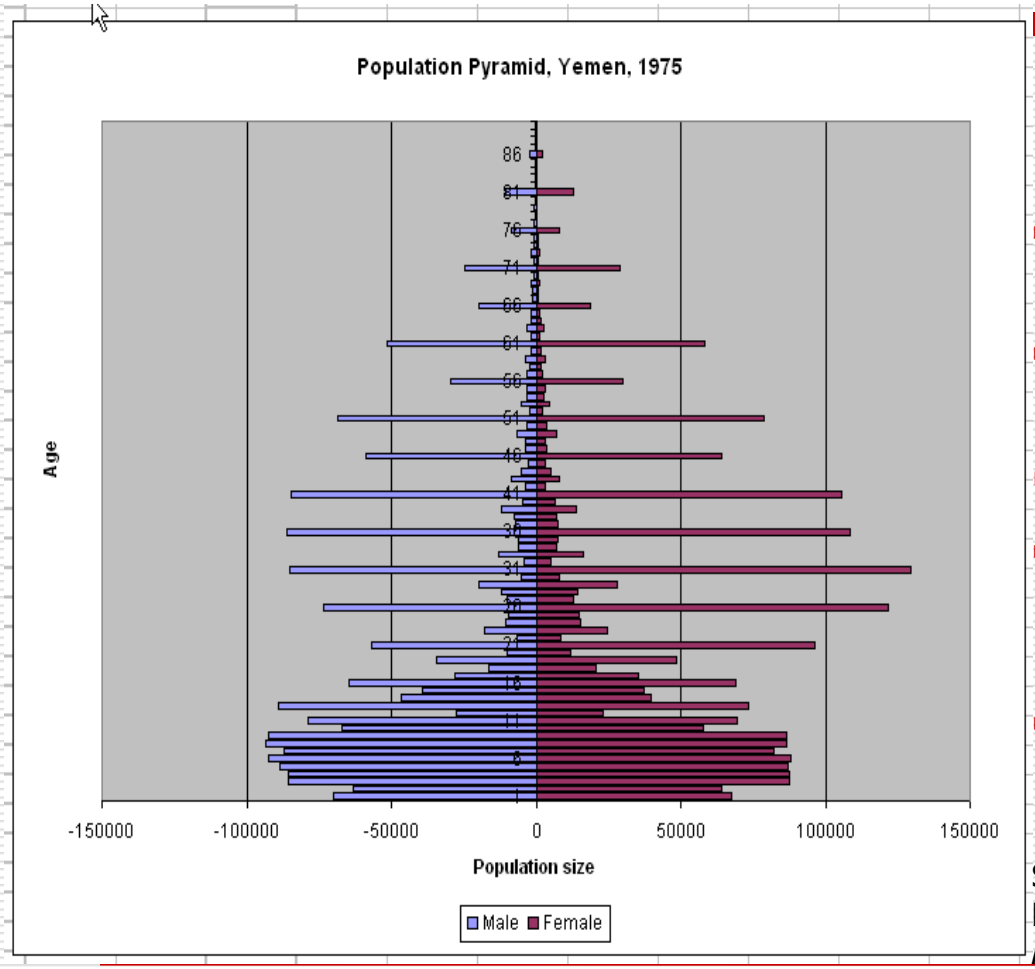


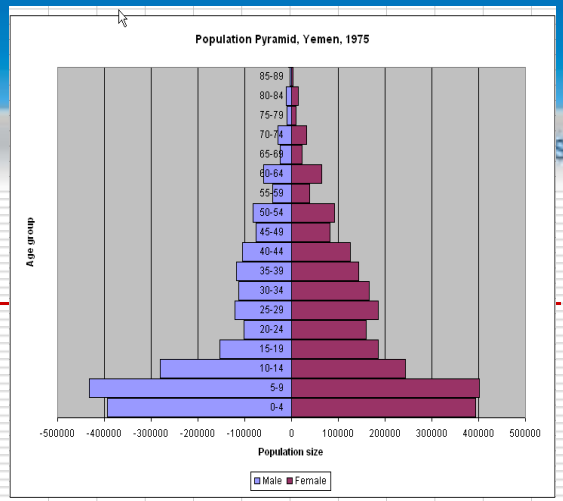
Source: Tabulated using data from *United Nations Demographic Yearbook*

Population pyramid (2) – low population growth



Population pyramid (3) - detecting errors

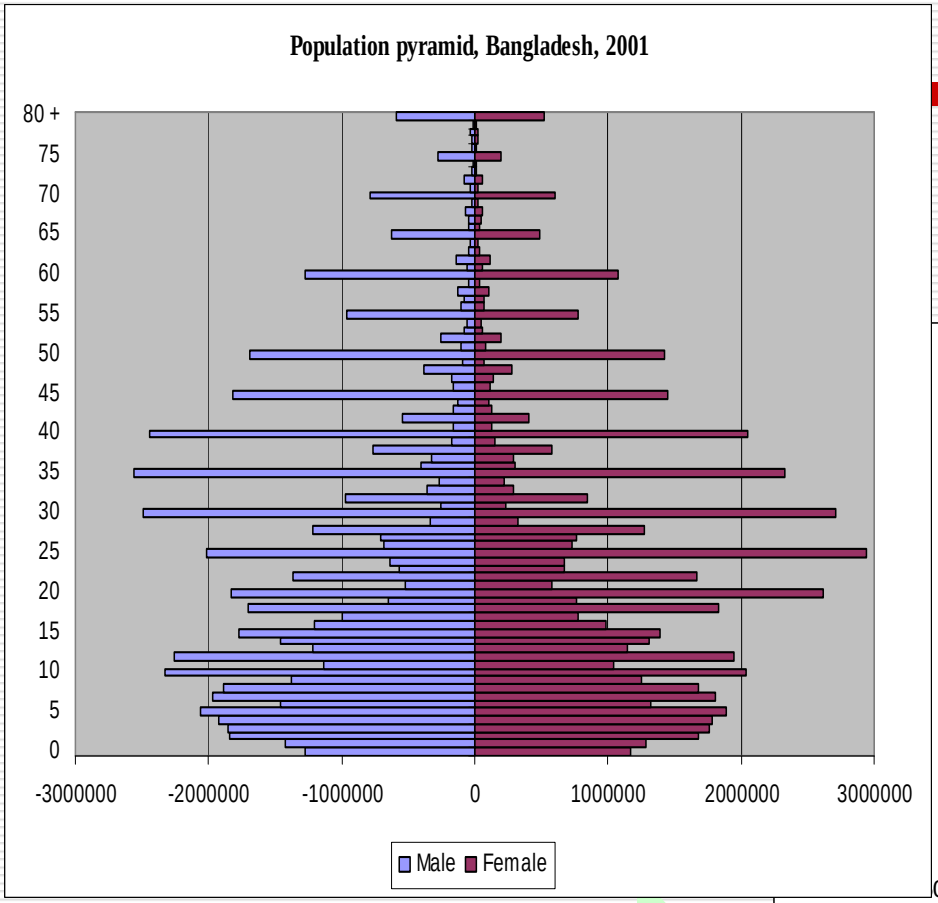




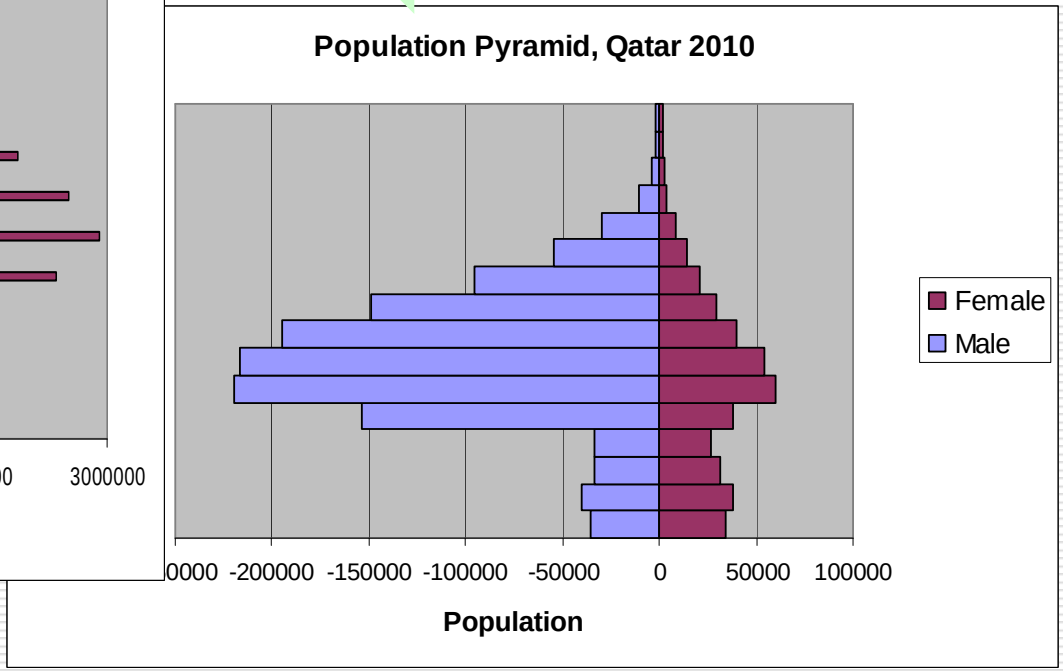
- Under enumeration of young children (< age 2)
- Age misreporting errors (heaping) among adults
- High fertility level
- Smaller population in 20-24 age group – extraordinary events in 1950-55?
- Smaller males relative to females in 20 – 44 - labor out-migration?

Source: Tabulated using data from U.S. Census Bureau, *Evaluating Censuses of Population and Housing*

Population pyramid (4)- detecting errors

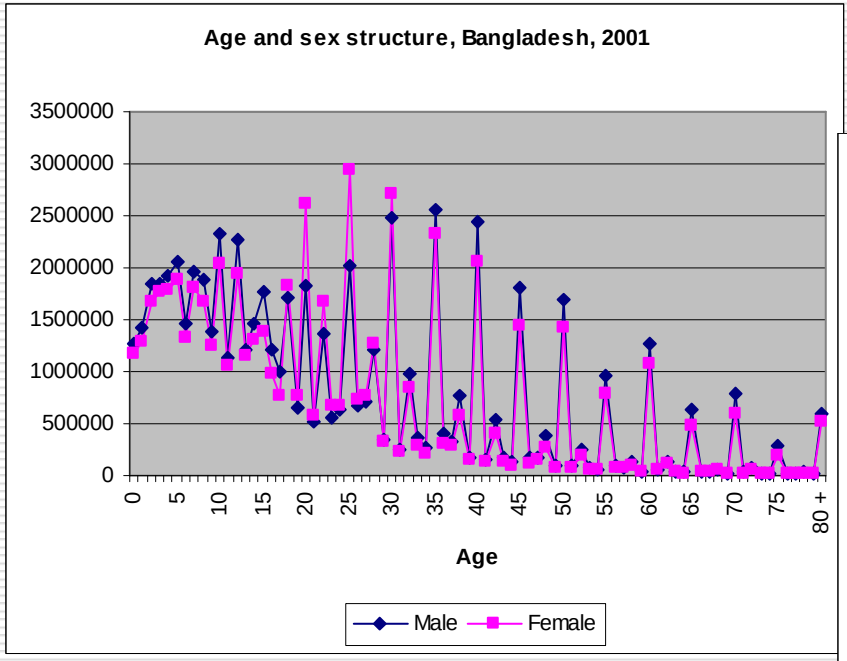


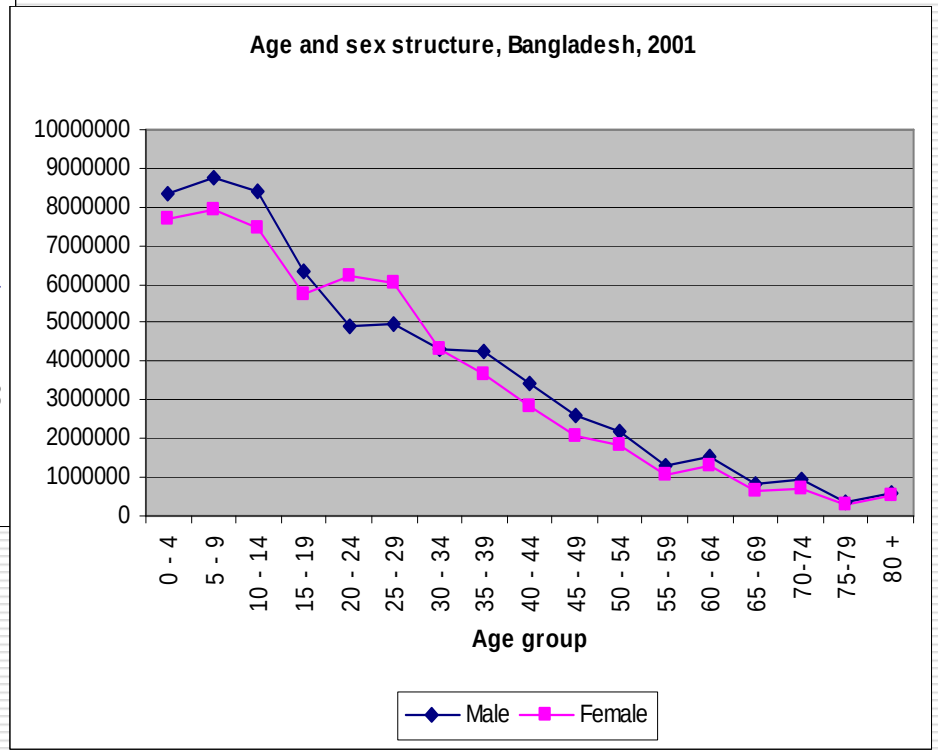
Age heaping



Labour in-migration

Population pyramid (5) - line instead of bars





Data source: Tabulated using data from *United Nations Demographic Yearbook*

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

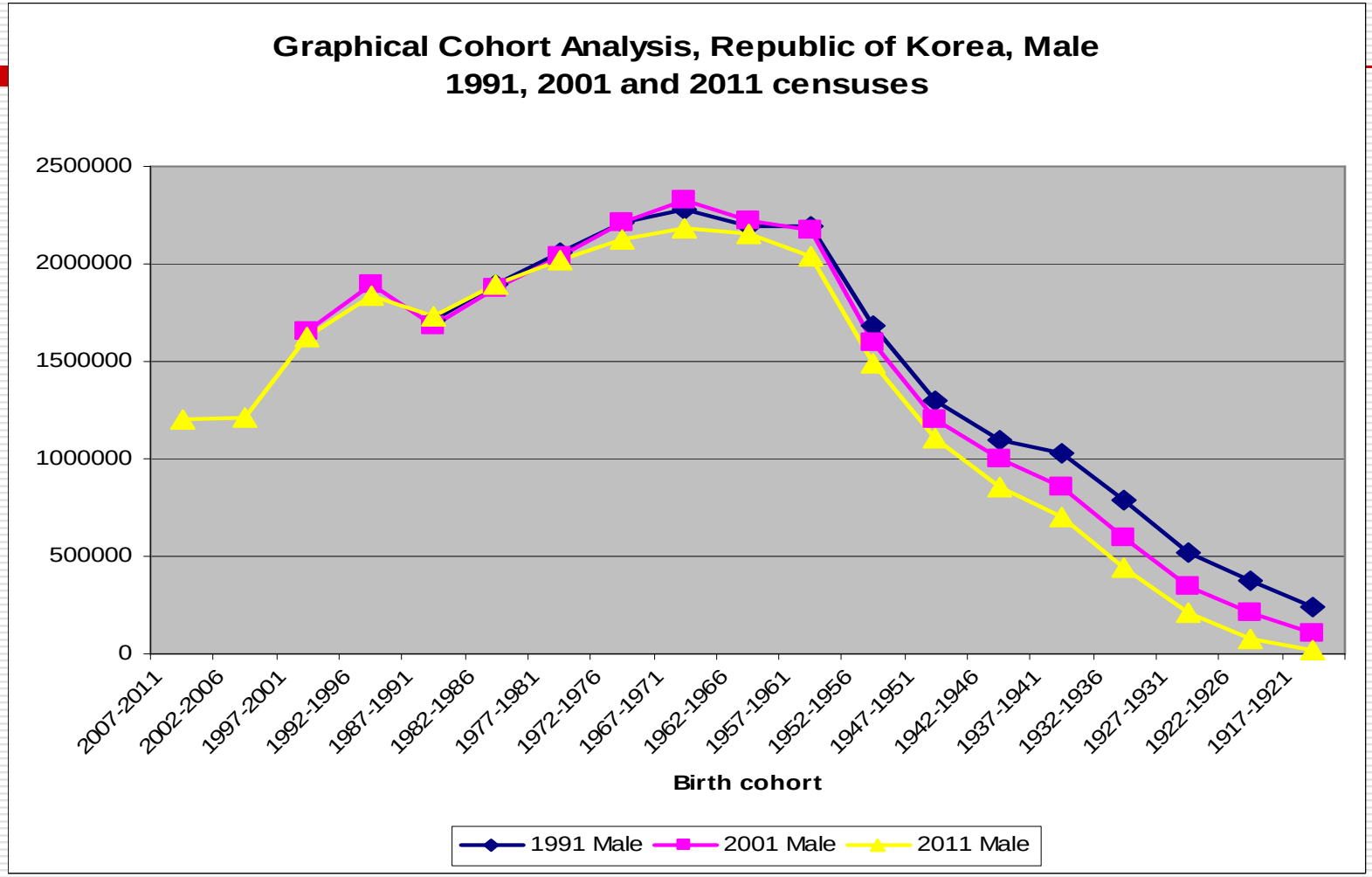
Graphical cohort analysis – Example (1)

Microsoft Excel - Population_five years age group_Rep. Korea_1991.xls

	C	D	E	F	G	H	I
4		1991	1991	2001	2001	2011	2011
5	Birth Cohort	Male	Female	Male	Female	Male	Female
6							
7	2007-2011					1197156	1127649
8	2002-2006					1213058	1125688
9	1997-2001			1655456	1510737	1627043	1480318
10	1992-1996			1897900	1672632	1839509	1628970
11	1987-1991	1711889	1526977	1680171	1507566	1733186	1541622
12	1982-1986	1892414	1752238	1873293	1748807	1890165	1751164
13	1977-1981	2052888	1922314	2039132	1939534	2020029	1925104
14	1972-1976	2215342	2081699	2209720	2115080	2120835	2023310
15	1967-1971	2282016	2147005	2325771	2240687	2187292	2120041
16	1962-1966	2189388	2074617	2218744	2086781	2155273	2070051
17	1957-1961	2187801	2071235	2174116	2086932	2040905	2013580
18	1952-1956	1679109	1594562	1597944	1556192	1492110	1518735
19	1947-1951	1301987	1243808	1204911	1188960	1105062	1175595
20	1942-1946	1095429	1060037	998979	1057538	857384	988875
21	1937-1941	1024861	1041324	858340	986591	705791	915967
22	1932-1936	788855	885534	593391	798132	446888	700349
23	1927-1931	515094	690840	345308	581841	206830	432044

- For this analysis we organize the data by birth cohort
- Exclude open-ended age category

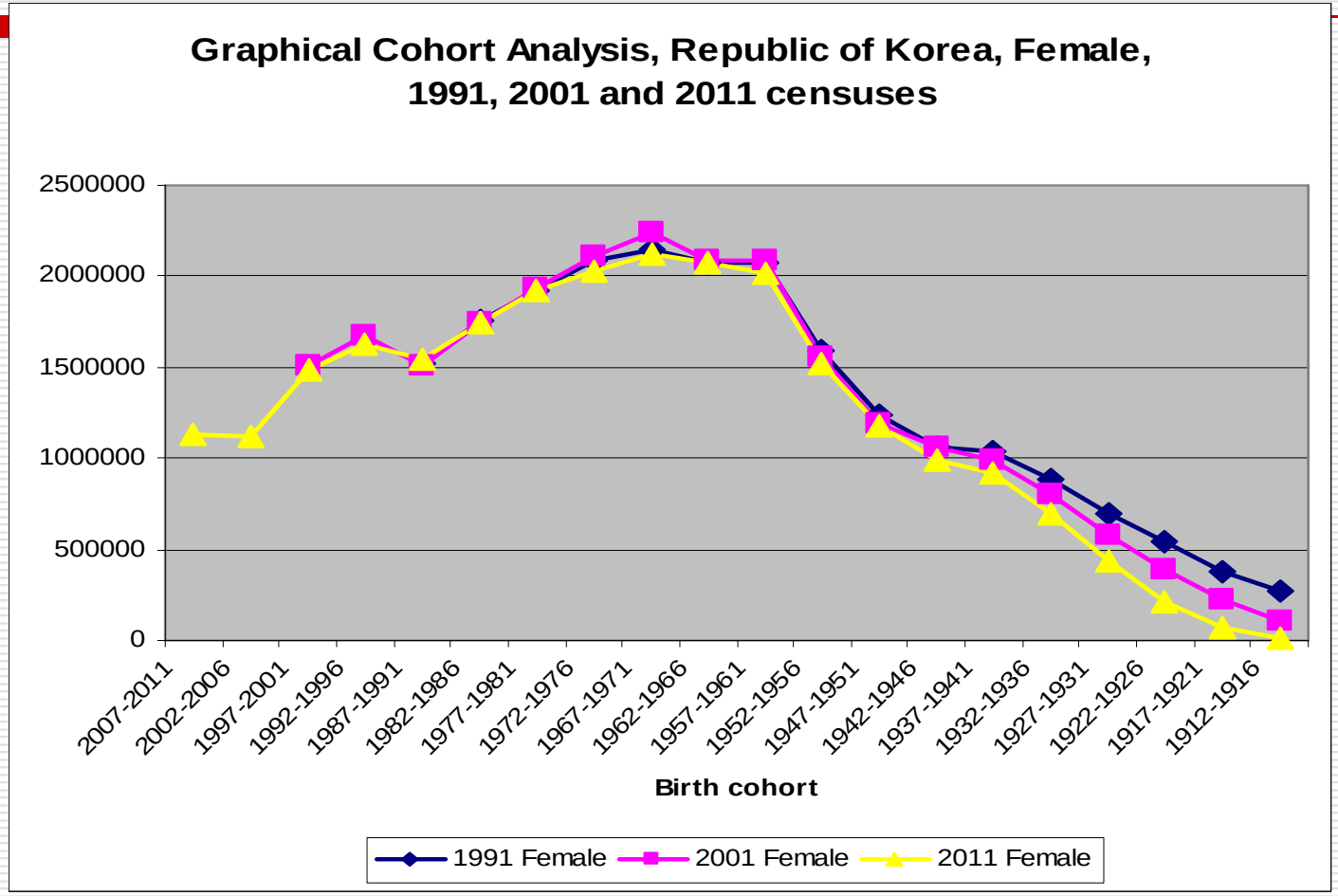
Graphical cohort analysis – Example (2)



Slide 14

1 HAOYI CHEN, 08/11/2012

Graphical cohort analysis – Example (2)



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

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

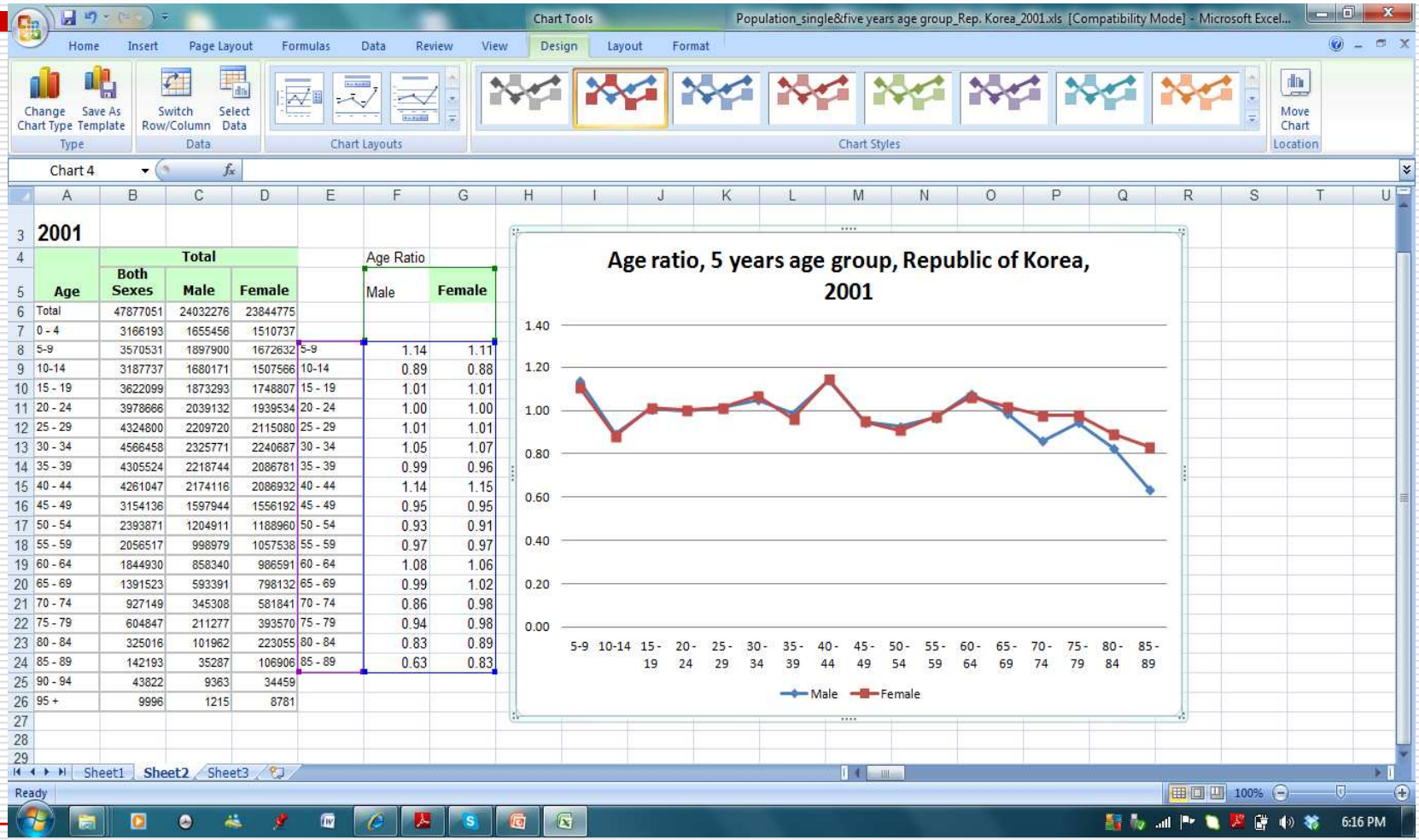
${}_5AR_x$ = The age ratio for the age group x to x+4

${}_5P_x$ =The enumerated population in the age category x to x+4

${}_5P_{x-5}$ = The enumerated population in the adjacent lower age category

${}_5P_{x+5}$ = The enumerated population in the adjacent higher age category

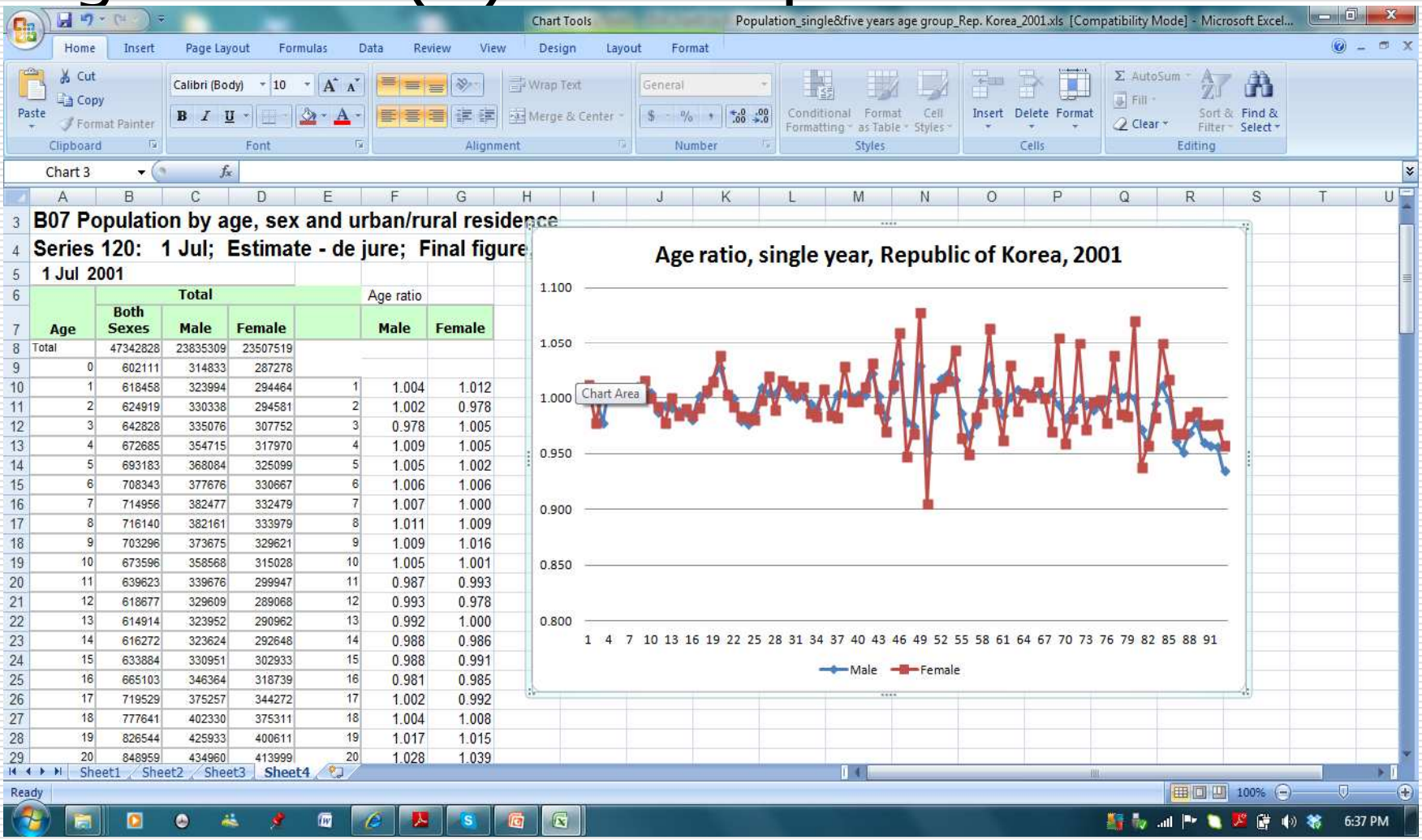
Age ratios (3) - example



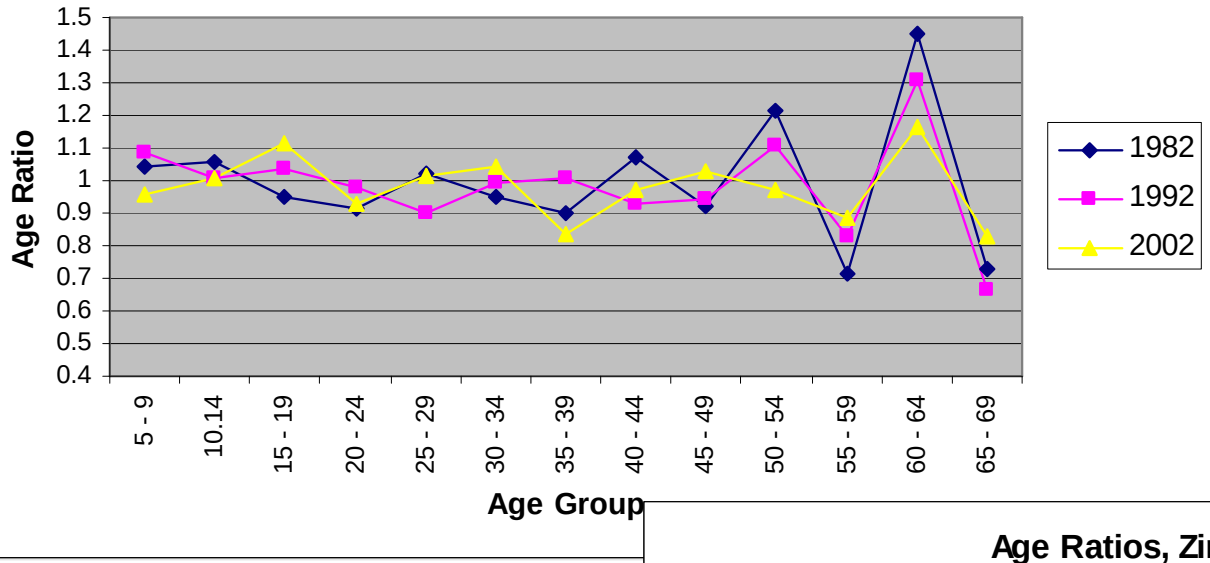
2001

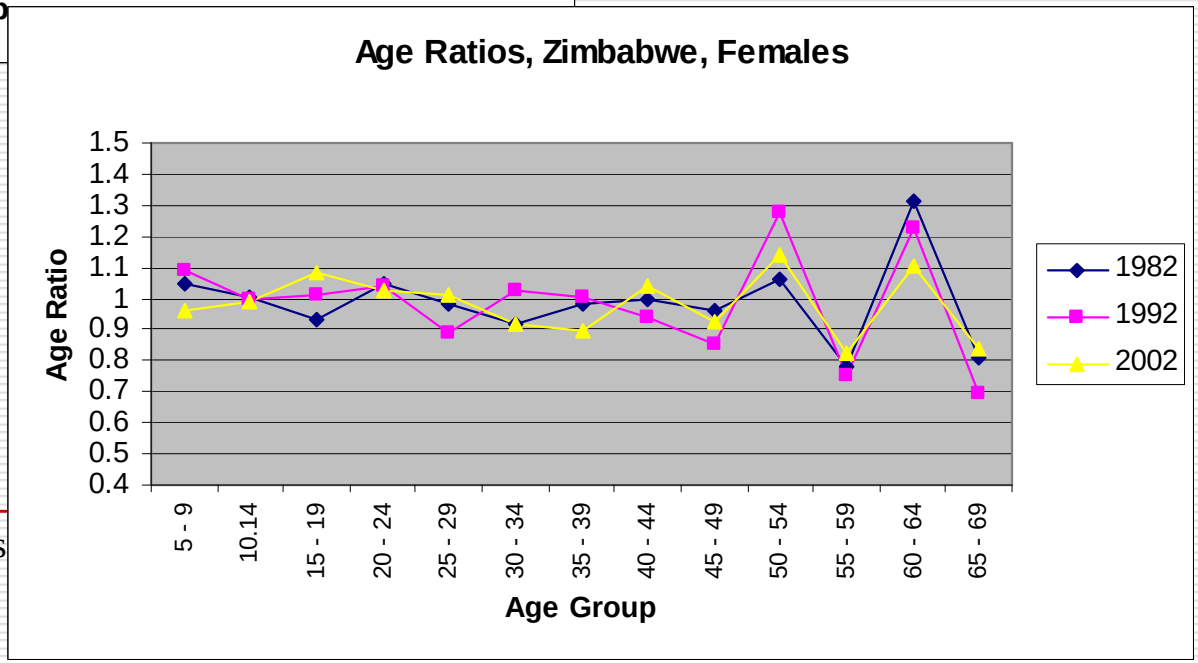
Age	Total			Age Ratio	
	Both Sexes	Male	Female	Male	Female
Total	47877051	24032276	23844775		
0 - 4	3166193	1655456	1510737		
5-9	3570531	1897900	1672632	1.14	1.11
10-14	3187737	1680171	1507566	0.89	0.88
15 - 19	3622099	1873293	1748807	1.01	1.01
20 - 24	3978666	2039132	1939534	1.00	1.00
25 - 29	4324800	2209720	2115080	1.01	1.01
30 - 34	4566458	2325771	2240687	1.05	1.07
35 - 39	4305524	2218744	2086781	0.99	0.96
40 - 44	4261047	2174116	2086932	1.14	1.15
45 - 49	3154136	1597944	1556192	0.95	0.95
50 - 54	2393871	1204911	1188960	0.93	0.91
55 - 59	2056517	998979	1057538	0.97	0.97
60 - 64	1844930	858340	986591	1.08	1.06
65 - 69	1391523	593391	798132	0.99	1.02
70 - 74	927149	345308	581841	0.86	0.98
75 - 79	604847	211277	393570	0.94	0.98
80 - 84	325016	101962	223055	0.83	0.89
85 - 89	142193	35287	106906	0.63	0.83
90 - 94	43822	9363	34459		
95 +	9996	1215	8781		

Age ratios (4) - example

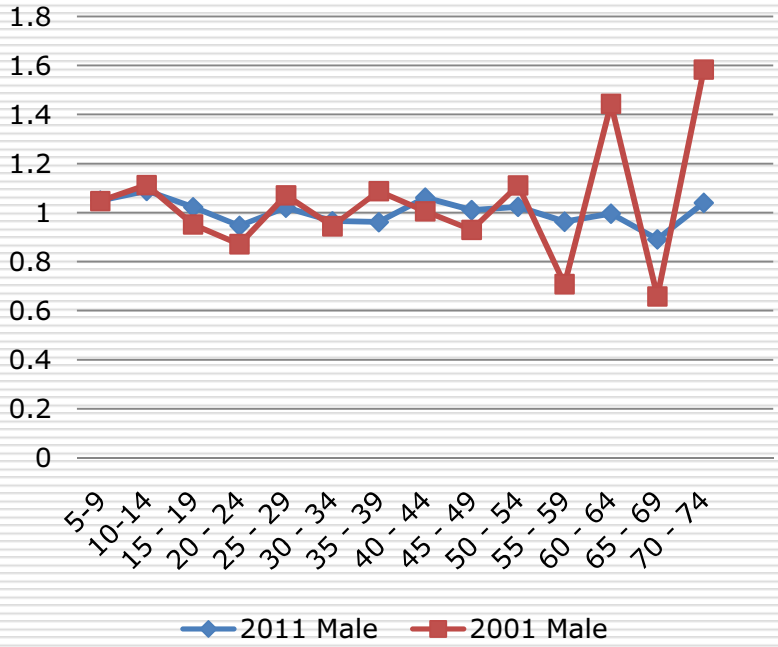


Age	Total: Both Sexes	Total: Male	Total: Female		Age ratio: Male	Age ratio: Female
Total	47342828	23835309	23507519			
0	602111	314833	287278			
1	618458	323994	294464	1	1.004	1.012
2	624919	330338	294581	2	1.002	0.978
3	642828	335076	307752	3	0.978	1.005
4	672685	354715	317970	4	1.009	1.005
5	693183	368084	325099	5	1.005	1.002
6	708343	377676	330667	6	1.006	1.006
7	714956	382477	332479	7	1.007	1.000
8	716140	382161	333979	8	1.011	1.009
9	703296	373675	329621	9	1.009	1.016
10	673596	358568	315028	10	1.005	1.001
11	639623	339676	299947	11	0.987	0.993
12	618677	329609	289068	12	0.993	0.978
13	614914	323952	290962	13	0.992	1.000
14	616272	323624	292648	14	0.988	0.986
15	633884	330951	302933	15	0.988	0.991
16	665103	346364	318739	16	0.981	0.985
17	719529	375257	344272	17	1.002	0.992
18	777641	402330	375311	18	1.004	1.008
19	826544	425933	400611	19	1.017	1.015
20	848959	434960	413999	20	1.028	1.039

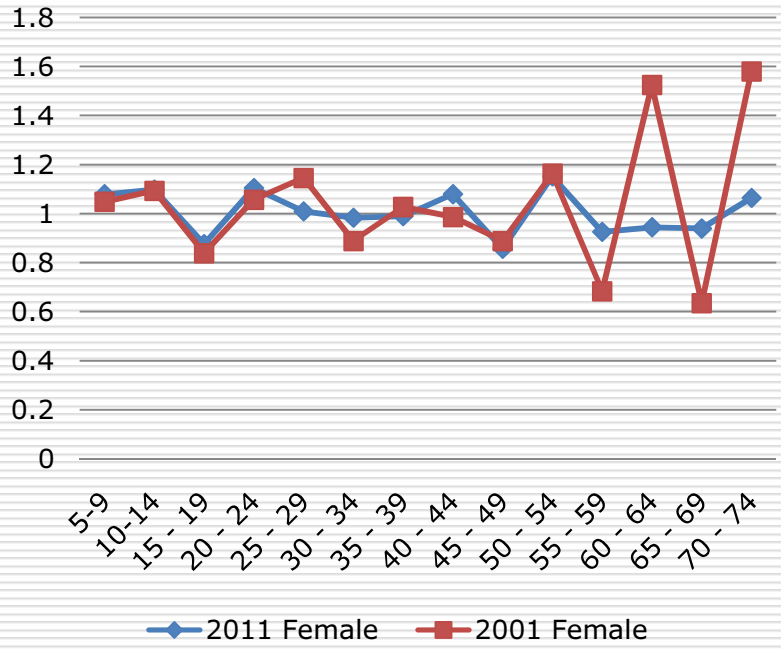

Age Ratios, Zimbabwe, Males
Age Ratio
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
5 - 9
10.14
15 - 19
20 - 24
25 - 29
30 - 34
35 - 39
40 - 44
45 - 49
50 - 54
55 - 59
60 - 64
65 - 69
Age Group
1982
1992
2002


Age Ratios, Zimbabwe, Females
Age Ratio
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
5 - 9
10.14
15 - 19
20 - 24
25 - 29
30 - 34
35 - 39
40 - 44
45 - 49
50 - 54
55 - 59
60 - 64
65 - 69
Age Group
1982
1992
2002

Age ratios, Bangladesh, Male



Age ratios, Bangladesh, Female



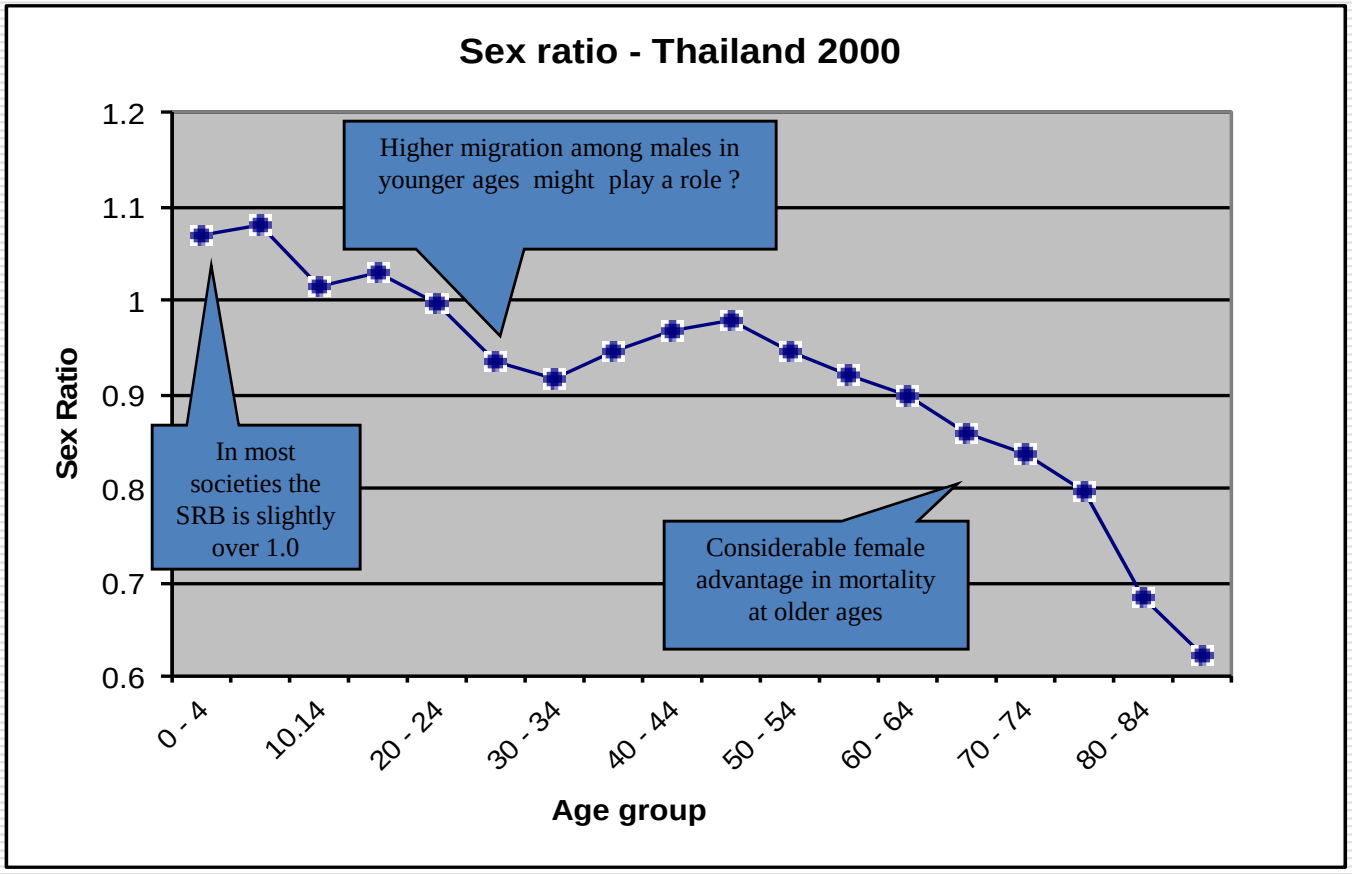
Sex ratios (1) - calculation

Sex Ratio = $_5M_x / {}_5F_x$

$_5M_x$ = Number of males enumerated in a specific age group

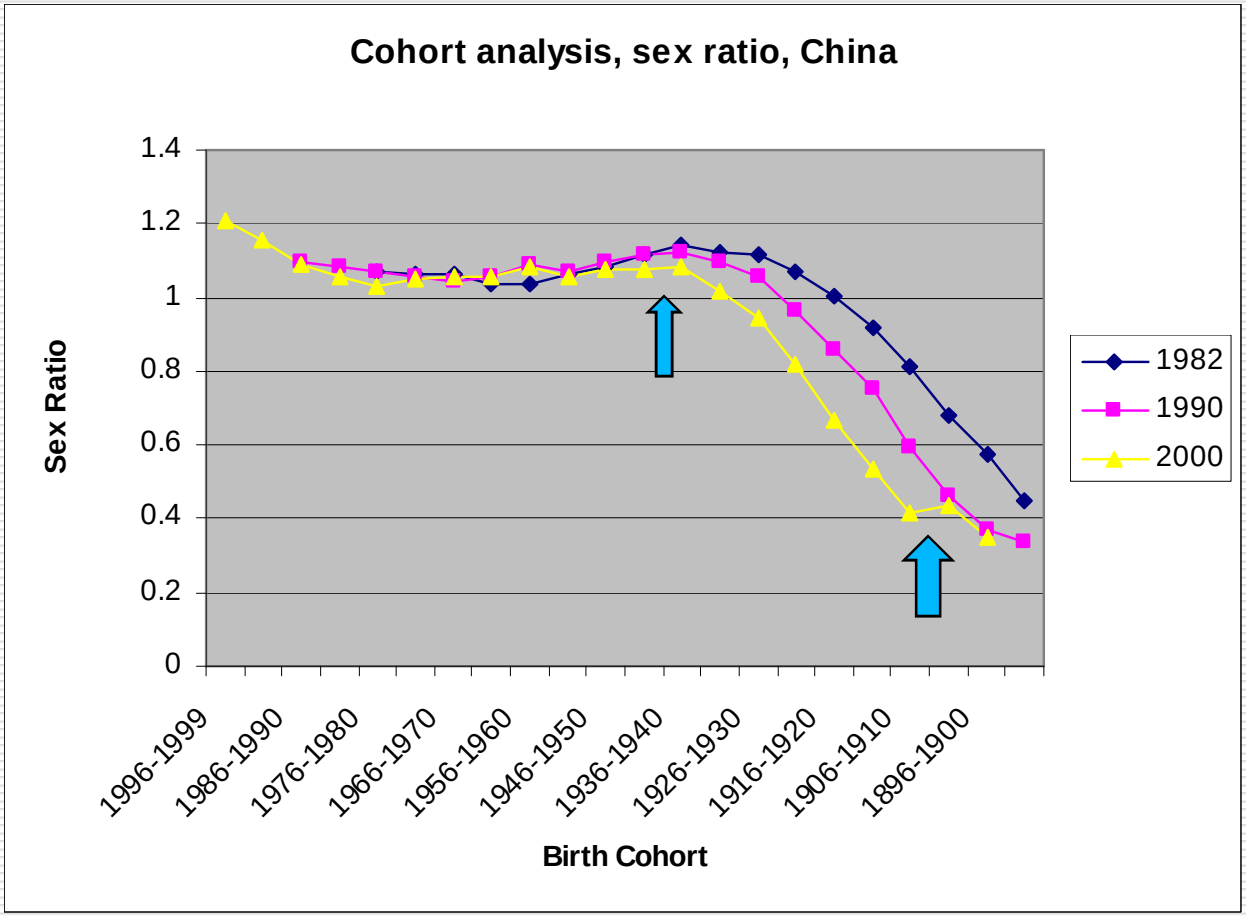
$_5F_x$ = Number of females enumerated in the same age group

Sex ratios (2) - plotting



Source: Tabulated using data from *United Nations Demographic Yearbook*

Sex ratios (3) – cohort analysis



Summary indices - Whipple`s Index

- Developed to reflect preference for or avoidance of a particular terminal digit or of each terminal digit
- The original Whipple`s index measures age heaping for the ages ending in 0 or 5
- It assumes a linear distribution of ages in each five year age range- linear decrease in the number of persons of each age within the age range considered
- The choice of the range 23 to 62 is standard, but largely arbitrary. In computing indexes of heaping, ages during childhood and old age 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, 1976, *Methods and Materials of Demography*

Whipple`s Index (2)

- Ranges between 100, *representing no preference for "0" or "5"* and 500, *indicating that only digits "0" and "5" were reported* in the census

- If heaping on terminal digits "0" and "5" is measured;

$$\frac{\sum (P_{25} + P_{30} + ... + P_{55} + P_{60})}{(1/5)\sum (P_{23} + P_{24} + ... + P_{60} + P_{61} + P_{62})} \times 100$$

- If the heaping on terminal digit "0" is measured;

$$\frac{P_{30} + P_{40} + P_{50} + P_{60}}{(1/10)\sum (P_{23} + P_{24} + ... + P_{60} + P_{61} + P_{62})} \times 100$$

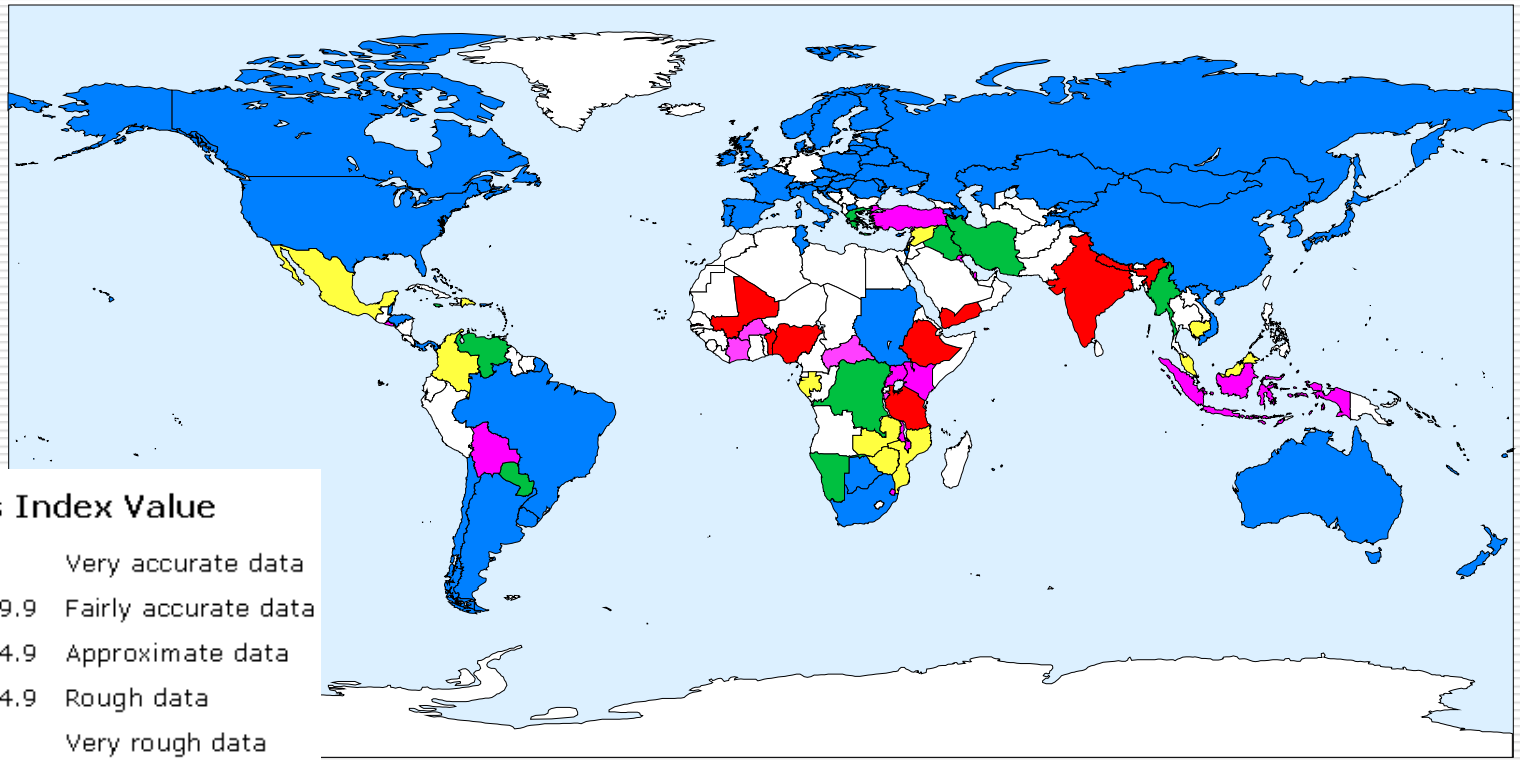
Whipple`s Index (3)

- The index can be summarized through the following categories:

	Value of Whipple's Index
• Highly accurate data	<= 105
• Fairly accurate data	105 – 109.9
• Approximate data	110 – 124.9
• Rough data	125 – 174.9
• Very rough data	>= 175

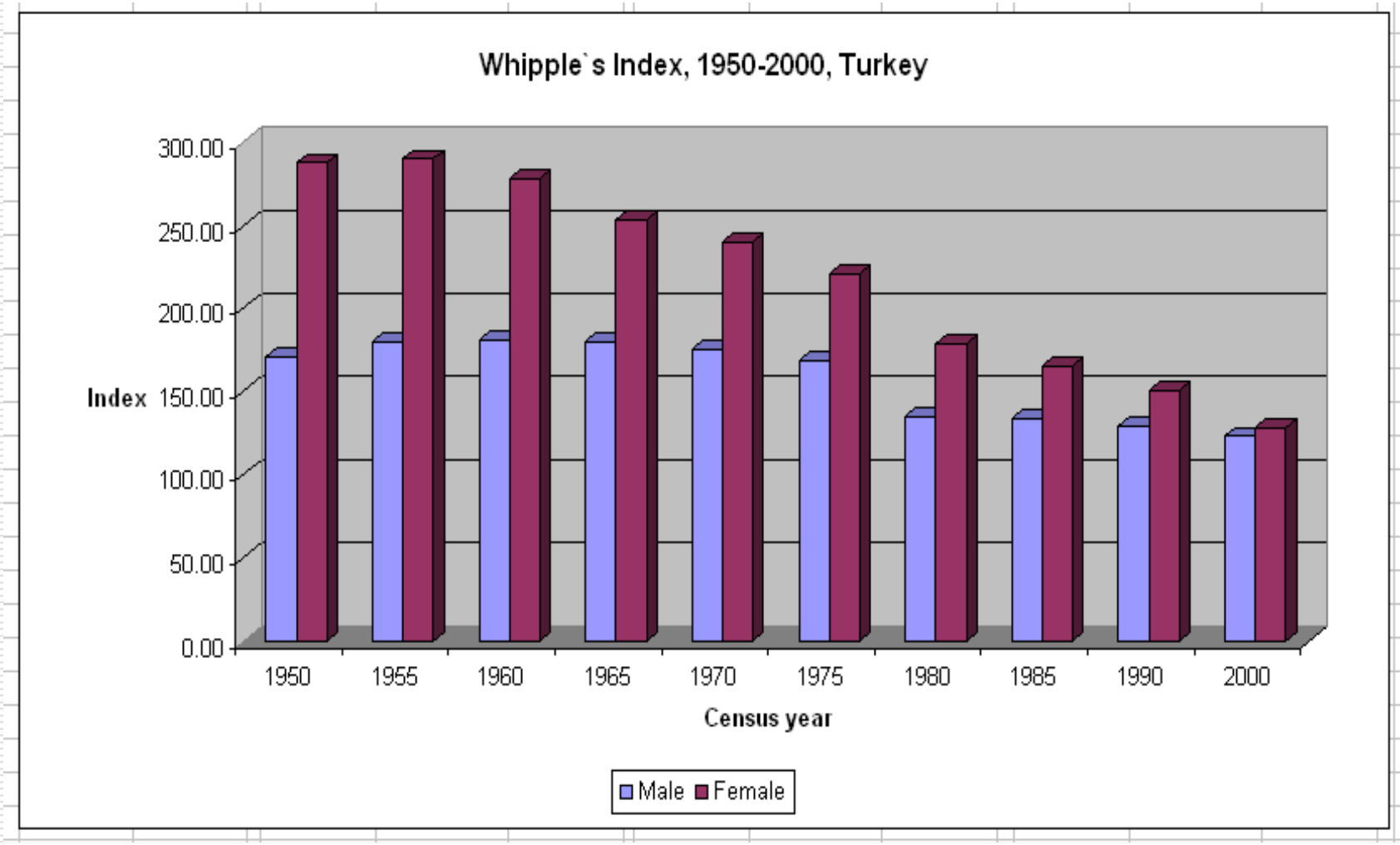
Whipple's index around the world

- Many of the countries that continue to have high Whipple's Index values are in Sub-Saharan Africa



Data source: *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



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
- It is based on the principle that in the absence of age heaping, the aggregate population of each age ending in one of the digits 0 to 9 should represent 10 % of population
- 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

Myers' Blended Index: Example

Microsoft Excel non-commercial use - Population_five years age grouo_Bangladesh_2011.xls [Compatibility Mode]

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Bangladesh 2001**												
2		**Male**	**Female**			**Myers` Blended Method-Bangladesh Census 2001-Females**							
13	10	2325193	2034385										
14	11	1137231	1048173			**Sum of population ages ending in terminal digit**		**Weights (given)**		**Blended Population**		**Deviation from 10%**	
15	12	2259700	1941453										
16	13	1211177	1147816		**Terminal digit**	**From 10+x**	**From 20+x**	**Column 1**	**Column 2**	**(1)x(3)+(2)x(4)**	**Row/Cell J28**	**abs(6)-10**	
17	14	1456134	1311117			*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	
18	15	1769894	**1392536**		**0**	12,521,437	10,487,052	1	9	106,904,905	28.46	18.46	
19	16	1209970	**985516**		**1**	2,154,416	1,106,243	2	8	13,158,776	3.50	6.50	
20	17	999020	**773996**		**2**	5,237,304	3,295,851	3	7	38,782,869	10.33	0.33	
21	18	1706287	**1825155**		**3**	2,345,847	1,198,031	4	6	16,571,574	4.41	5.59	
22	19	646732	**767374**		**4**	2,400,234	1,089,117	5	5	17,446,755	4.65	5.35	
23	20	1827496	2623507		**5**	9,576,307	8,183,771	6	4	90,192,926	24.01	14.01	
24	21	516934	577905		**6**	2,251,745	1,266,229	7	3	19,560,902	5.21	4.79	
25	22	1370308	1669967		**7**	2,102,798	1,328,802	8	2	19,479,988	5.19	4.81	
26	23	564204	669623		**8**	4,142,463	2,317,308	9	1	39,599,475	10.54	0.54	
27	24	638938	676473		**9**	1,389,238	621,864	10	0	13,892,380	3.70	6.30	
28	25	2019173	**2938082**		**Total**					375,590,550		66.69	
29	26	680305	**724225**										
30	27	708991	**764711**					Summary index of age preference = Total/2				33.35	
31	28	1211640	**1275930**										
32	29	338216	**320741**										
33	30	2489499	**2706528**										
34	31	259229	**237026**										
35	32	971399	847528										
36	33	359182	287775										
37	34	262665	220888										

Sheet1 Sheet2 Sheet3 age ratio2 Sheet5 age ratio Myers

11:02 PM

Conclusion: Uses and limitations

- Assessment of the age and sex structure of the population enumerated in a census is typically the first step taken in evaluating a census 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 census 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 census counts for particular segments of the population. Estimates of coverage error from other sources often are required to verify these observations.

References

- Shryock and Siegel, 1976, *Methods and Materials of Demography*
- IUSSP Tools for Demographic Estimation (in progress)

  http://demographicestimation.iussp.org/