Documentation for

Children per Woman (Total Fertility Rate)

for countries and territories

ROUGH DRAFT FOR COMMENTS

No quotes please

Gapminder Documentation constitutes work in stepwise progress.
We welcome all sorts of comments, corrections and suggestions through e-mail to the author.

Gapminder Documentation 008
When together with revisions: Version 2
Uploaded: 2010-01-29
Published by: The Gapminder Foundation,
Sweden, Stockholm, 2009

Author: Ferenc Ajus, with inputs from Mattias Lindgren
E-mail: mattias.lindgren (at) Gapminder.org
Gapminder is a non-profit foundation
promoting sustainable global development
by increased use and understanding of statistics.
www.gapminder.org

0. Introduction

This is the documentation for the Gapminder compilation of Children per Woman (Total Fertility Rate). The data is used in the interactive graph *Gapminder World* available at:

www.gapminder.org/world

This written documentation is accompanied with an excel-file, which includes both the detailed meta-data, as well as the actual figures. This excel-file will henceforth be referred to as the "excel-file". The excel file, as well as this documentation, are available at:

www.gapminder.org/downloads/documentation/#gd008

The main purpose of the data is to produce graphical presentations that display the magnitude of income disparities in the world over time. Therefore, we have also included very rough estimates for countries and territories for which reliable data was not available. These estimates can only be taken as an indication of the order of magnitude for the indicator. Furthermore, we have not been able to make sure that every single observation is based on the best estimates available. Hence we discourage the use of this data set for statistical analysis and advise those who require more exact data to investigate the available data more carefully and look for additional sources, when appropriate.

There are at least two purposes for including very rough estimates:

a) It gives a starting point for a discussion of the data. Unrealistic estimates become more apparent if they are displayed in a format that many can access and is easy to assess. This, we hope, might elicit comments and criticism from others, which would help us to improve the data step by step.

b) We want to display the broad patterns of development of the world. Hence, the observations before 1900 (or even 1950) are less aimed at describing the relative position of individual countries. Rather, we want to display our best guesses for the development of the world as a whole.

0.1 The countries included in the dataset

For a discussion on what countries and territories we try to cover, and how we try to handle border changes and the like, see the document "Countries and Territories in Gapminder World". The basic principle, however, is to get estimates for the geographical areas corresponding to the present borders.

To be succinct, we will hereafter jointly refer to all countries and other types of geographical entities and territories as "countries", irrespectively of their statehood status. The inclusion of any area in this data set does, in no way, imply a stated opinion of Gapminder on the legal status of the area.

0.2 Overview

Our dataset contains historic total fertility rate (TFR) data for 195 countries in the world, for the 1800-2008 time period, within their current borders. The main source of the data is the UN fertility dataset, which covers the period between 1950 and 2008. In addition, it has been complemented with data collected from statistical sources and various published articles. For almost all countries, the data now cover the period when fertility changed from high (e.g. 6 children per woman) to low levels (e.g. 2 children). For less developed countries the UN dataset was enough for this, because the transition happened after 1950. For more developed countries we could successfully extend the dataset backwards in time, to see what the pre-transition level of fertility was. Based on this, we could also estimate the 1800 level of fertility for all countries.

The dataset may be used freely, but users are warned that it contains both good quality observations and rough estimates. It is easy to distinguish those, however. Data quality is indicated in the dataset in field "qual." for each record. Data with quality indicator 4-5 are estimates, and should not be used for research. Quality 1-2 are taken from a statistical source, published article or book, which is reliable in our view. Data quality 3 contains *crude birth rates* from a reliable source, but as TFRs were not available in the source, the authors estimated that using a simple re-scaling conversion. You may also see the source of each data record in the dataset, cross-referenced to a list of literature used.

0.3 This report

We have compiled a historic Total Fertility Rate (TFR)[1] dataset for 195 countries of the world. The purpose of this report is to explain the methods applied in some detail. The report will aim at:

1. Explaining the approach (some theoretical background)
2. Showing main sources of fertility data we have used
3. Explaining estimation and extrapolation methods used in case of missing or limited data
4. Explaining the structure of the dataset collected
5. Showing research background and references

Appendices include data sources, examples and the data themselves in graphs by region.

[1] The total fertility rate or TFR is the sum of age-specific fertility rates. It can be interpreted as the number of children an „average woman" would have in the given population. In other words TFR gives the average number of children that would be born to a woman over her lifetime if (1) she were to experience the statistically observed yearly age-specific fertility rates through her lifetime, and (2) she were to survive from birth through the end of her reproductive life.

1. The approach

Good coverage of the fertility transition

The aim of the project was to collect a dataset that may be used to show the transition from high to low fertility in an interactive moving graph. This objective was achieved reasonably well, partly using reliable sources, and partly using estimates. The general tendency of demographic transition has been captured in a way that is in line with the current knowledge in historical demography. (see also: Caldwell, 2006; Chesnais, 1992)

Approach: cover the transition period with data, assume stable fertility before that

The data collection was guided by the principle that as much as possible, estimates should be avoided, and instead of searching for an estimation model, it is better to search for more data. This approach builds on the knowledge what we have about fertility transition: as the explanations of fertility transition and pre-transition fertility levels are controversial, we do not have a model that would predict fertility with reasonable accuracy from other data. The transition from high to low fertility was usually quick; it went through within a few decades. If the period of transition is not covered with data, then it is very difficult to make good estimates. However, pre-transition fertility (or in Henry's terms "natural fertility") was more stable, and a few data points should be enough to be able to make a rough estimate of that.

Luckily, most of the time data availability coincides with modernization and modernization coincides with fertility transition. This means that with some exceptions, we have data for the period of fertility transition, and we only have problems before that, which causes less inaccuracy, because "natural fertility" is less volatile.

Keeping this in mind, during the second stage of data collection we identified those countries where fertility transition started before 1950 (not being covered by UN data), and concentrated the effort on those.

Limitation 1.: We know little about the level of "natural fertility"

While the data are suitable for showing the general tendencies of the transition from high to low fertility, the fluctuation and level of pre-transition fertility remains unknown for most countries, and it was out of scope for this project.

Limitation 2.: Data quality: use of estimates

This dataset should only be used for research after filtering out the estimates. In its current state it includes both rough estimates and statistical data. This is indicated in the field "qual." The general rule applied was as follows:

1 TFR data taken from statistical source or published article

2 mainly used for the UN data (1950-2008), which includes some estimates

3 crude birth rates (or Princeton If indexes) were taken from statistical source or published article, but TFR was estimated by the author (using a simple conversion)

4 no country-wide data were available, but a published estimate was found. This estimate is usually based on historical demographic research for a smaller area.

5 estimate (or guess) by the author

In addition, when there is some inconsistency in the time series, or between sources, which cast doubt on the data quality, we indicated this in the column "Possible error in data", and lowered data quality to 4, even if the data were from an official source. This has only been done in a few obviously strange cases. In most cases we assumed that data from published sources are reliable, and awarded quality indicator 1-3 without undertaking detailed source critique.

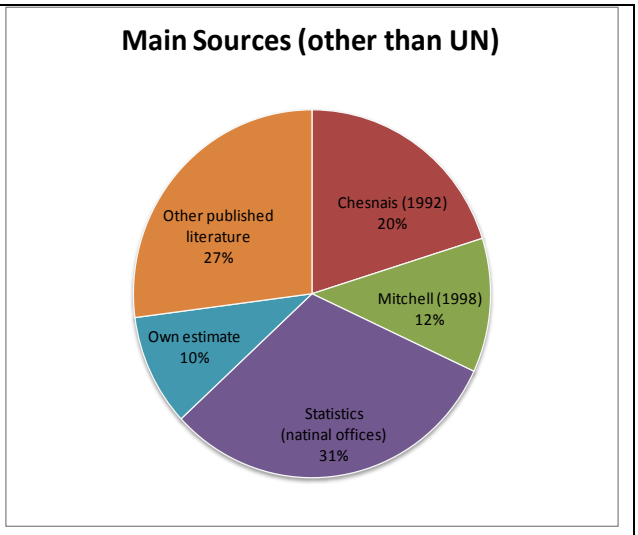
This means that only TFR data with quality 1-2 and crude rates with quality 1-3 may be used for research.

2. Sources and results of data collection

Between 2008 and 1950 the UN fertility dataset was used. Our work mainly concerned finding data sources before 1950, with a special attention to those countries, where the transition started earlier than that. The following table shows the number of records collected by main sources:

Table 1. Count of records in the dataset by source

Sources of records	
UN dataset (2008)	11230
Own collection	3561
Total data rows	14791
"Own collection" in more detail:	
Chesnais (1992)	717
Mitchell (1998)	420
Statistics	1110
Own estimates	350
Other published literature	964
Total data rows, own coll.	3464



The "other" category includes the following:

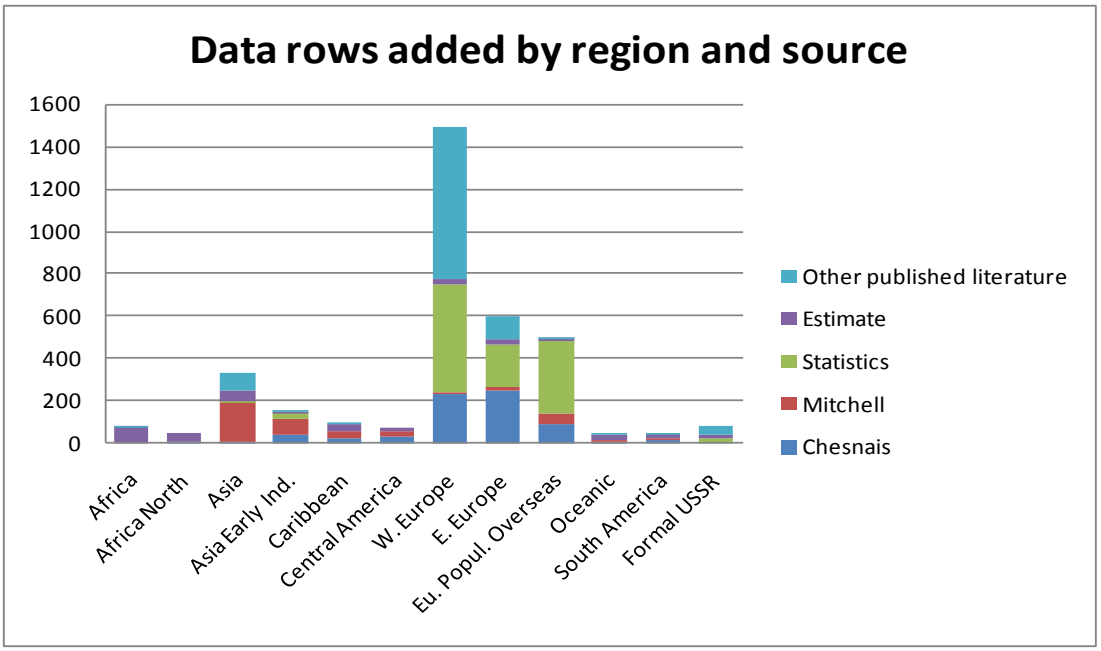
- books about Historic Fertility and fertility transition
- search for articles about individual countries
- dataset provided by Tomas Sobotka[2]

[2] Tomas Sobotka has written several articles about fertility, using TFR rates for Europe. He is a participant in the Human Fertility Database project and he expressed some interest to link to our database later.

- Princeton European Fertility Project (I_f indexes were converted)

The full list of sources can be found in the dataset on sheet "Sources" and also as Appendix 2 in this report.

The data collected may also be shown by region. I have used a demographically meaningful regionalization, as it was proposed by Caldwell (2006).



As you may observe on the graph, for the more developed countries much more data were available, but this is also in line with our approach sated above. You may see the details of these categories in graphs in the appendix, showing fertility transition by country within each region.

3. Estimation and extrapolation methods used

The main challenge was to estimate the TFR values for 1800. This is the starting point of interpolation for the software, but for most countries we had no data so far back in time.

Three different estimation methods were used, depending on the situation:

1. Assuming that fertility observed in 1950 was natural, so it was the same earlier. For countries where fertility was still high (above 5.5) in 1950, we attempted to find the level of natural fertility based on the data that we have. This involved calculating the average of the years 1950-55, or if there were earlier data, trying to find a "plateau" around which the data tend to fluctuate. Essentially, we assumed that the 1950-55 level of fertility is natural fertility, and in 1800 fertility was similar. *In case of countries where we had no data between 1800 and 1925, we have repeated this estimate for 1925*, to show, that we assume high natural fertility for the whole period. This was also needed to avoid a situation when data are automatically extrapolated between a high point in 1800 and a lower point in 1950. While the assumption of high

natural fertility in general for these countries is in line with the literature, the year 1925 was selected in an arbitrary way, with this technical purpose in mind.

2. Going back in time with the data until "natural fertility" is identified. For the rest of the countries as the transition started earlier, the UN data were not enough. In most cases the effort to cover fertility transition with at least some data was successful.
    a. Extrapolation backwards: In most cases we could find data for earlier years, and could look for a "plateau" in the data, to see what the pre-transition fertility level was. We have used this level as an estimate for fertility in 1800.
    b. Estimate from literature: For some countries we found an expert estimate of "natural fertility" in the literature, or some article which enabled us to estimate the pre-transition level of TFR. (Example: Vishnevskij (2006)includes specific estimates for pre-transition TFR in ex-USSR countries; while Saito (2006) includes information on fertility level in two provinces of India, from which we could estimate the rough level of TFR for India as a whole.
    c. Analogy with neighbor: mainly in the case of small countries which have not been independent earlier, and we were lacking a separate source, the data from a neighboring country were copied.

3. Exceptions: There were only a few countries, where, neither (1.) nor (2.) was feasible. In these cases we applied regional averages, or data from neighboring countries. The full list of these countries is shown in Table 2. In case of a few more countries we had to make our estimates based on 1-3 observations, looking at the regional data at the same time. These are shown in Table 3. [3]

Table 2.
Countries where the initial phase of the decline is not covered with data
(estimates for 1800 were based on regional average + random factor, or based on data of neighboring country)

Country	Besides the UN, how many observations we had	Estimated TFR (1800)	UN first TFR (1950)
Central African Republic	1	6.508	5.4
Gabon	2	6.500	4.1
Gambia, The	1	6.512	5.4
Sierra Leone	1	6.518	5.5
Hong Kong, China	1	5.47	4.5

[3] Tables 2 and 3 contain all the exceptions. The estimate for the rest of the countries is more reliable, as they either had their transition during the period covered by UN data, or we could find 6 or more observations before 1950.

Macao, China	2	5.51	5.3
Bahamas, The	2	5.90	4.0
Grenada	1	5.81	5.2
St. Lucia	2	5.53	5.3
Guam	1	6.37	5.4
New Caledonia	1	6.41	5.1
Reunion	1	6.51	5.5
French Guiana	1	6.27	5.1

(criteria used: TFR 1950< 5.5 and obs. without UN < 3)

Table 3.
Countries where the initial phase of the decline is scarcely covered with data
(estimates for 1800 were based on the few existing country observations before 1950, using also regional comparisons)

Country	Besides the UN, how many observations we had	Estimated TFR (1800)	UN first TFR (1950)
Channel Islands	3	5.07	2.1
Montenegro	5	5.90	3.5
Slovenia	5	4.93	3.1
Bosnia and Herzegovina	5	5.91	5.1
Moldova	4	6.39	3.5
Belarus	4	7.00	3.4
Georgia	4	7.80	3.1
Kazakhstan	3	6.58	5.0
Kyrgyz Republic	3	6.60	4.3
Ukraine	4	7.50	2.7

(criteria used: TFR 1950< 5.5 and obs. without UN < 6 but >2)

Estimating TFR from crude birth rates

For historic times (before 1950) TFR is often not available, because we do not know the age-specific birth rates. In these cases we had to estimate TFR from the crude birth rates. Here we used the method suggested by Bogue (1993). This textbook includes an empirical table of equivalent values between four different fertility measures. Among others, this includes conversion rates between Crude Birth Rate (CBR), and Total Fertility Rates (TFR). This empirical conversion was based on regression using UN data. The conversion is relatively

exact, as the correlation is 0,989 between CBR and TFR. We did not use this table directly, but instead we have implemented the same regression method on our dataset. (see an example in Appendix 3.) If possible the conversion rate was determined by country, using this method, but when there were not sufficient TFR data, we have applied a general conversion factor calculated from the UN dataset. (TFR=0,134*CBR)

The following table gives the conversion rates for all countries, where a country-specific conversion was applied.

Table 3.
Slope (conversion rate between crude birth rate and TFR) calculated for individual countries

Austria	0.1381
Belgium	0.1461
Denmark	0.1350
France	0.1500
Germany	0.1346
Hungary	0.1220
Japan	0.1478
Netherlands	0.1460
Norway	0.1439
Portugal	0.1300
Spain	0.1323
Sweden	0.1418
Tunisia	0.1490
United Kingdom	0.1358
United States	0.1193

In all cases where TFR was calculated from crude rates, both were included in the dataset in separate columns. As the conversion is only approximately correct, we have lowered data quality to "3" in these cases.[4]

TFR estimates from Princeton data (for Europe around 1900)

Researchers within the framework of the European Fertility Project have used Princeton fertility indexes (I_f). These indexes compare the fertility of a given region with the fertility of the Hutterites, a population where no birth control was exercised due to religious reasons. The data have been published, and they cover Europe on a county level, supplying 4 observations between 1880 and 1940. These data can be converted to TFR with some

[4] The formula calculating TFR in column C may include a general conversion rate (taken from the cell named „Conv"= 0,134, or a country-specific conversion rate, shown in column H). Both kinds of conversion rates are shown in column H, but the general one is only shown as a text, the formula uses the "Conv" cell instead. This is done to ensure that it would be enough to change the „Conv" field on sheet „Key" to alter the general conversion rate everywhere.

accuracy, because they include some information about age distribution. (e.g.: We know the TFR of the Hutterites, and we know that $I_f = 0.6$ means that fertility is 60% of the Hutterites, from this TFR can be determined.) Sardon (1996) suggests 12,44 as a conversion key. We have applied this conversion to fill the data gap for a few European countries, like Baltics and ex-Jugoslav states.[5]

4. The structure of dataset

I have followed the structure of the existing Gapminder life expectancy dataset. Please look at sheet "DATA" and "Sources".

Rows:
In sheet DATA, rows are yearly observations, or estimates. The number of rows is shown in Table 1.

Columns:
"Crude birth rate column (CBR)": we show both crude birth rate and TFR in cases where TFR was converted by us using a rate. In other cases we usually do not fill the CBR column. There is a quality indicator from 1-5. The conversion key is shown in each line where a conversion from CBR to TFR was applied. The note and remark fields were used to show observation-specific comments, like basis of estimation, information about sources or secondary sources, reference years (e.g. when the source refers to a five year interval).

Sheets:
I use a separate sheet to name major sources, and then cross-reference that to the list ("Chesnais" means that in Chesnais(1992) book there was a crude rate, from which I have estimated the TFR). In addition, a web page address is entered directly into the DATA sheet if possible.

You may filter for Spain as an example; it has both original and estimated data. (see: Appendix 3.)

5. Theoretical background

The main purpose of literature review was to provide input for estimating the TFR in developing countries in the 19th century.

Chesnais (1992) – was the second main source of data after UN. It includes a lot of useful data, and analysis of various forms of Fertility Transition around the world.

[5] The Princeton dataset is available on-line at : http://opr.princeton.edu/archive/pefp/.

Mitchell (1998): Is probably the best book on historical statistics; it comes in three large volumes. It included some fertility (crude birth rate) data, which overlapped with Chesnais (1992), but sometimes complemented it.

Dyson (1985) – includes data about the onset of fertility decline in the World.

Sardon (1996) – gives the method and multiplier to convert Cole's indexes to TFR

Caldwell (2006) – includes a two page comprehensive table which estimates the date of onset and speed of fertility decline in all major regions of the world, and includes data by country as well. The country coverage is broader than in Chesnais, but the data are less exact. We have not used this table directly, because our dataset is more detailed. However, the classification of the countries was very informative, and we have implemented this in order to check our data. For example Caldwell created a separate category for "European Population Overseas", for countries like the US, Australia and Chile, where fertility transition started about as early as in Europe, while most non-European countries started the decline much later, after the 1950s.

Appendix

Appendix 1. – Literature (used in this report)

1. **Chesnais**, Jean-Claude (1992): The Demographic Transition : Stages, Patterns, and Economic Implications. A Longitudinal Study of Sixty-Seven Countries Covering the Period 1720-1984; *Oxford : Clarendon Press , 1992*

2. **Mitchell**, B.R. (1998): International Historical Statistics Macmillan, Volumes I-III. UK 1998

3. **Sardon**, Jean-Paul (1996): Coale's Indices, Comparative Indices, Mean Generation, Total Fertility Rate and Components, *Population: An English Selection, Vol. 8 (1996), pp. 252-257*

4. **Caldwell, John C.** (2006) : Demographic Transition Theory; Chapter 10.: Regional Paths to Fertility Transition, Dordrecht, The Netherlands: Springer. pp. 418. ISBN 1-4020-4373-2.

5. **Bogue,** Donald J. (et al) (1993): Empirical interrelationships among standard fertility measures. Readings in population research methodology. Volume 3. Fertility research, edited by Donald J. Bogue, Eduardo E. Arriaga, Douglas L. Anderton, George W. Rumsey. Chicago, Illinois, Social Development Center, 1993. :11-51 - 11-53. Published for the United Nations Population Fund [UNFPA]

6. **Dyson**, Tim and **Murphy**, Mike (1985): The Onset of Fertility Transition. Population and Development Review, Vol. 11, No. 3 (Sep., 1985), pp. 399-440

7. **Watkins**, Suzan Cotts (1987): The Fertility Transition: Europe and the Third World Compared: Sociological Forum, Vol. 2, No. 4, Special Issue: Demography as an Interdiscipline (Autumn, 1987), pp. 645-673

8. **Wrigley**, E. A. (1978) "Fertility strategy for the individual and the group." In C. Tilly (ed.), Historical Studies of Changing Fertility: 135-154. Princeton, NJ: Princeton University Press.

Appendix 2. – Data sources

Data sources and references (explanations of the abbreviations used in the data sheet)	
Abbreviations:	**Explanations:**
Main sources	
STAT	National Statistical Office (of the given country)
UN Pop. 2008	UN World Population Prospects 2008 dataset (contains smoothened data end estimations)
Chesnais	Chesnais, Jean-Claude. (1992) The Demographic Transition : Stages, Patterns, and Economic Implications. A Longitudinal Study of Sixty-Seven Countries Covering the Period 1720-1984 Jean-Claude Chesnais Oxford : Clarendon Press , 1992
Mitchell	Mitchell, B.R. (1998): International Historical Statistics Macmillan, Volumes I-III. UK 1998
own estimate	Estimate by Gapminder, as explained in the documentation.
+	"+" sign used when there are two different sources (like one for crude rate, and a different one for TFR)
Further sources	
Banens	Banens, M.; Bassino J. and Egretaud E. (1998): "Estimating population and labour force in Vietnam under French rule (1900-1954) Institute of economic research Hitotubashi university (Discussion Paper No. D98-7) 1998, July from table 22.
Barclay	Barclay, G.W.; Coale, A.J.; Stoto, M.A.; Trussell J. T. (1976); A Reassessment of the Demography of Traditional Rural China Population Index, Vol. 42, No. 4 (Oct., 1976), pp. 606-635 Office of Population Research
Caldwell	New Data on Nuptiality and Fertility in China Author(s): John C. Caldwell and K. Srinivasan Source: Population and Development Review, Vol. 10, No. 1 (Mar., 1984), pp. 71-79 Published by: Population Council

Calot	Gérard Calot, Two centuries of Swiss demographic history. Graphic album of the 1860–2050 periods. Table F_05_01 SWITZERLAND, 1932-1939 AGE-SPECIFIC FERTILITY RATE by age of mother reached during the calendar year of the birth Source to quote : SWISS FEDERAL STATISTICAL OFFICE/OBSERVATOIRE DEMOGRAPHIQUE EUROPEEN, 1998
Castles	Castles, Lance, 1975. Sources for the Population History of Northern Sumatra. Masyarakat Indonesia Tahun Ke 2/2 pp 189-208 comment(Data from one region in Northern Sumatera(Batak)) Based on baptism records data was in the following book: 東南アジア人口民族誌 "坪内良博", 1986
Corruccini	Corruccini, R.S., Handler, J.S. (1983): Plantation Slave Life in Barbados: A Physical Anthropological Analysis Journal of Interdisciplinary History, Vol. 14, No. 1 (Summer, 1983), pp. 65-90
Cz_stat	Czech Statitistical Office (from Sobotka)
Festy	Festy, Partick. (1979): "La fécondité des pays occidentaux de 1870 à 1970." Travaux et Documents No. 85. Paris: INED – PUF.
Freedman	Freedman R., Ming-Cheng Chang, Te-Hsiung Sun: Taiwan's Transition from High Fertility to Below-Replacement Levels Studies in Family Planning, Vol. 25, No. 6 (Nov. - Dec., 1994), pp. 317-331 Published by: Population Council
Gentile	Gentile, Michael: "Population Geography Perspectives on the Central Asian Republics", Swedish Institute for Futures Studies Stockholm 2005, ISBN 91-89655-74-5
Heuser	Heuser, R. L. 1976. Fertility tables for birth cohorts by colour. U. S. Department of Health, Education, and Welfare. DHEW Publication No. (HRA)76-1152. Schoen, R. 2003. "Insights from parity status life tables for the 20the century U.S." Working Paper 03-08, Population Research Institute, The Pennsylvania State University.
Hollerbach	Fertility determinants in Cuba by: Paula E. Hollerbach, Sergio Diaz-Briquets (National Academy Press, 1983, Washington)
Insee	TABLEAU 44 - TAUX DE FÉCONDITÉ GÉNÉRALE PAR ÂGE DE LA MÈRE Nombre d'enfants nés vivants pour 10 000 femmes de chaque âge INSEE 2005
JP	Population Statistics of Japan 2008, Table 4.3
Keyfitz	Keyfitz, Nathan. (1968): World Population : Analysis of Vital Data Nathan Keyfitz ; Flieger, Wilhelm Chicago : University of Chicago Press , 1968 . - X, 672

Kollega	Kollega T. I. (ed): Hungary in the 20st century - Magyarország a XX. Században (Babits , 1996)
Kucera	Vladimír a Kučera, Milan. Vývoj obyvatelstva českých zemí v XIX. století. In.:Statistika a demografie I., ČSAV, Praha, 1959.
Kumo	Kazuhiro KUMO, Takako MORINAGA, Yoshisada SHIDA. (2007): Long-Term Population Statistics for Russia, 1867-2002 RRC Working Paper Series No. 2 December 2007
Ludwig	Ludwig, Armin K. (1985): "Brazil: A handbook of Historical Statistics " ISBN 0816181888
LU-MADD	Lund University Macroeconomic and Demographic Database
Lutz	Lutz, Wolfgang; Scherbov, Sergei; Volkov, Andrei G. (1994) Demographic trends and patterns in the Soviet Union before 1991 Routledge, New York, 1994
Mexico	Klein, 'Familia y fertilidad en Amatenago, Chiapas, 1785— 1816.' HM, 36/2 (1986), 273—86.
Noriko	Noriko, Tsuta. (2001) "Birth regime of Modern Japan : Event history analysis in 18th and 19th Century Two Northeastern Villages" KESDP(Keio Economic Society Discussion Paper series) No.00-9 (in Japanese)
Pantelides	Pantelides E.A.: In: Expert Group Meeting on Completing the Fertility Transition, New York, 11-14 March 2002, [compiled by] United Nations. Department of Economic and Social Affairs. Population Division. New York, New York, United Nations, Department of Economic and Social Affairs, Population Division, 2004. :333-342.
Phil Census	Census of the Philippines Islands: 1903 Volume1 to 4 Director: Gen. J. P. Sanger, USA United States Bureau of the Census, Washington, 1905
Princeton	Princeton European Fertility Project. If indexes were shown in the dataset, and converted using 12,44 as a conversion key. (Sardon, 1996)
Rédei	Rédei (1960): A születések és halálozások alakulása KJK 1960
Rele	Rele JR Source: ASIA-PACIFIC POPULATION JOURNAL. 1988 Jun;3(2):29-54
Rothenbacher	Rothenbacher, F. (2002): "The European Population 1850-1945". CD-Rom. Basingstoke: Palgrave 2002
Saito	Saito, Osamu (2006): Pre-Transition Fertility in Asia: A Comparative-historical Approach; Journal of International Economic Studies (2006), No.20, 1–17
Sardon	Table 2 in: Sardon, J.-P. 1991. "Generation replacement in Europe since 1900." Population: An English selection 3(1990): 15-32.

Seers	Seers, Dudley (1957) A Fertility Survey in the Maltese Islands Population Studies, Vol. 10, No. 3 (Mar., 1957), pp. 211-228 Population Investigation Committee
Smith&Ng	Smith & Ng , 1882 The components of Population Change in Nineteenth-century Southeast Asia: Village Data from the Philippines. Population Studies, 36-2 comment((Data from one region (Nagla prefecture))) Based on Baptism record data was in the following book 東南アジア人口民族誌 "坪内良博", 1986
Sobotka set	Tomas Sobotka shared his dataset with Gapminder. Data taken from his file are indicated in column "Remark", but as in other cases, the source reference used by Sobotka is copied and indicated as "Source"
Sprocha	SPROCHA, Branislav - TIŠLIAR, Pavol. 2008. Plodnost a celková reprodukcia obyvatelstva Slovenska v rokoch 1919 - 1937. STIMUL: Bratislava
Vishnevskij	Vishnevskij (ed). (2006): Demographic modernisation of Russia: 1900-2000 Демографическая модернизация России: 1900-2000 Под редакцией Анатолия Вишневского Серия "Новая история" Москва, Новое издательство, 2006, 601 страница

Appendix 3. – Dataset example for Spain

CTRY	Year	Total Fertility Rate (TFR)	Crude Birth Rate (CBR)	Princeton If index	Qual.	Source	Conversion key	Region
Spain	1800	5.13			5	own estimate		EU
Spain	1850	5.13			5	own estimate repeated (assumed natural fertility at least till this date)		EU
Spain	1858	4.66	35.2		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1859	4.75	35.9		3	Rothenbacher (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1860	4.86	36.7		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1861	5.16	39		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1862	5.09	38.5		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1863	5.00	37.8		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1864	5.19	39.2		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1865	5.11	38.6		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1866	5.07	38.3		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1867	5.09	38.5		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1868	4.72	35.7		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1869	4.90	37		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1870	4.84	36.6		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1878	4.78	36.1		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1879	4.74	35.8		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1880	4.70	35.5		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1881	4.91	37.1		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1882	4.79	36.2		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1883	4.71	35.6		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1884	4.86	36.7		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1885	4.80	36.3		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1886	4.86	36.7		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1887	4.78	36.1		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1888	4.82	36.4		3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU

Spain	1889	4.82	36.4	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1890	4.55	34.4	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1891	4.67	35.3	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1892	4.71	35.6	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1893	4.71	35.6	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1894	4.60	34.8	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1895	4.63	35	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1896	4.75	35.9	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1897	4.51	34.1	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1898	4.41	33.3	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1899	4.53	34.2	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1900	4.49	33.9	3	Chesnais (CBR), own estimate TFR, using conv. key	0.132	EU
Spain	1901	4.71	35	1	Sardon		EU
Spain	1905	4.66	35.1	1	Sardon		EU
Spain	1910	4.43	32.7	1	Sardon		EU
Spain	1915	4.22	30.9	1	Sardon		EU
Spain	1920	4.14	29.5	1	Sardon		EU
Spain	1922	4.02	30.5	1	Sardon		EU
Spain	1923	4.02	30.5	1	Sardon		EU
Spain	1924	3.92	29.7	1	Sardon		EU
Spain	1925	3.82	29.1	1	Sardon		EU
Spain	1926	3.87	29.6	1	Sardon		EU
Spain	1927	3.70	28.1	1	Sardon		EU
Spain	1928	3.80	29.1	1	Sardon		EU
Spain	1929	3.69	28.3	1	Sardon		EU
Spain	1930	3.68	28.3	1	Sardon		EU
Spain	1931	3.58	27.6	1	Sardon		EU
Spain	1932	3.64	28.2	1	Sardon		EU
Spain	1933	3.59	27.7	1	Sardon		EU
Spain	1934	3.38	26.2	1	Sardon		EU
Spain	1935	3.31	25.7	1	Sardon		EU
Spain	1936	3.18	24.7	1	Sardon		EU
Spain	1937	2.89	22.6	1	Sardon		EU
Spain	1938	2.56	20.0	1	Sardon		EU
Spain	1939	2.12	16.5	1	Sardon		EU
Spain	1940	3.09	24.4	1	Sardon		EU
Spain	1941	2.47	19.5	1	Sardon		EU
Spain	1942	2.53	20.2	1	Sardon		EU
Spain	1943	2.88	22.9	1	Sardon		EU

Spain	1944	2.84	22.5	1	Sardon	EU
Spain	1945	2.91	23.1	1	Sardon	EU
Spain	1946	2.70		1	Sardon	EU
Spain	1947	2.67		1	Sardon	EU
Spain	1948	2.88		1	Sardon	EU
Spain	1949	2.68		1	Sardon	EU
Spain	1950	2.49	19.8	2	UN Pop. 2008	EU
Spain	1951	2.51	19.9	2	UN Pop. 2008	EU
Spain	1952	2.55	20.2	2	UN Pop. 2008	EU
Spain	1953	2.59	20.5	2	UN Pop. 2008	EU
Spain	1954	2.63	20.7	2	UN Pop. 2008	EU
Spain	1955	2.67	20.9	2	UN Pop. 2008	EU
Spain	1956	2.70	21.1	2	UN Pop. 2008	EU
Spain	1957	2.74	21.3	2	UN Pop. 2008	EU
Spain	1958	2.77	21.4	2	UN Pop. 2008	EU
Spain	1959	2.80	21.5	2	UN Pop. 2008	EU
Spain	1960	2.83	21.5	2	UN Pop. 2008	EU
Spain	1961	2.86	21.5	2	UN Pop. 2008	EU
Spain	1962	2.88	21.5	2	UN Pop. 2008	EU
Spain	1963	2.90	21.4	2	UN Pop. 2008	EU
Spain	1964	2.91	21.2	2	UN Pop. 2008	EU
Spain	1965	2.92	21.0	2	UN Pop. 2008	EU
Spain	1966	2.93	20.8	2	UN Pop. 2008	EU
Spain	1967	2.93	20.6	2	UN Pop. 2008	EU
Spain	1968	2.93	20.4	2	UN Pop. 2008	EU
Spain	1969	2.93	20.2	2	UN Pop. 2008	EU
Spain	1970	2.92	20.0	2	UN Pop. 2008	EU
Spain	1971	2.90	19.8	2	UN Pop. 2008	EU
Spain	1972	2.88	19.6	2	UN Pop. 2008	EU
Spain	1973	2.85	19.4	2	UN Pop. 2008	EU
Spain	1974	2.82	19.1	2	UN Pop. 2008	EU
Spain	1975	2.76	18.7	2	UN Pop. 2008	EU
Spain	1976	2.69	18.3	2	UN Pop. 2008	EU
Spain	1977	2.60	17.7	2	UN Pop. 2008	EU
Spain	1978	2.49	17.0	2	UN Pop. 2008	EU
Spain	1979	2.37	16.3	2	UN Pop. 2008	EU
Spain	1980	2.24	15.5	2	UN Pop. 2008	EU
Spain	1981	2.11	14.6	2	UN Pop. 2008	EU
Spain	1982	1.97	13.8	2	UN Pop. 2008	EU
Spain	1983	1.85	13.1	2	UN Pop. 2008	EU
Spain	1984	1.73	12.4	2	UN Pop. 2008	EU
Spain	1985	1.64	11.8	2	UN Pop. 2008	EU
Spain	1986	1.55	11.3	2	UN Pop. 2008	EU
Spain	1987	1.48	10.9	2	UN Pop. 2008	EU
Spain	1988	1.43	10.6	2	UN Pop. 2008	EU

Spain	1989	1.38	10.3	2	UN Pop. 2008	EU
Spain	1990	1.34	10.1	2	UN Pop. 2008	EU
Spain	1991	1.30	9.9	2	UN Pop. 2008	EU
Spain	1992	1.27	9.8	2	UN Pop. 2008	EU
Spain	1993	1.24	9.6	2	UN Pop. 2008	EU
Spain	1994	1.22	9.5	2	UN Pop. 2008	EU
Spain	1995	1.20	9.4	2	UN Pop. 2008	EU
Spain	1996	1.19	9.4	2	UN Pop. 2008	EU
Spain	1997	1.19	9.4	2	UN Pop. 2008	EU
Spain	1998	1.19	9.4	2	UN Pop. 2008	EU
Spain	1999	1.20	9.6	2	UN Pop. 2008	EU
Spain	2000	1.22	9.7	2	UN Pop. 2008	EU
Spain	2001	1.24	9.9	2	UN Pop. 2008	EU
Spain	2002	1.27	10.1	2	UN Pop. 2008	EU
Spain	2003	1.30	10.4	2	UN Pop. 2008	EU
Spain	2004	1.33	10.6	2	UN Pop. 2008	EU
Spain	2005	1.36	10.7	2	UN Pop. 2008	EU
Spain	2006	1.39	10.9	2	UN Pop. 2008	EU

TFR was missing before 1900, but I found crude rates in Chesnais. An overlapping period 1901 -1934 was used to calculate the slope assuming 0 intercept.

Linear regression	
r	0.99976758
r squared	0.999535214
Observations	18

	Coeff.	*Std. error*	*t value*	*p-value*
Intercept	0			
X slope	0.132314132	0.000692005	191.2039565	8.95521E-30

The above value (0,132314) was used as a conversion factor to get TFR before 1900 from crude rates. This assumes stable age distribution.

For 1800 and 1825 I have included my own estimate, as even crude rates are missing. This estimate is based on the idea that we already have natural fertility in 1850, I only assumed a slightly higher level.