Documentation for

Total Population

for countries and territories

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

Gapminder Documentation 003
Version 1
Uploaded: 2008-10-08
Published by: The Gapminder Foundation,
Sweden, Stockholm, 2008

Author: 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

1. Introduction

This documentation is for the Gapminder compilation "Total Population". Total population is, by default, used for the size of the bubbles in Gapminder World: www.gapminder.org/world

An indicator for the data quality of the individual observations in "Total Population" is also available in Gapminder World called "Data quality – Total Population". Four spreadsheets accompany this documentation:

- The "Population data (xls)": includes all the available observations (the data is not always yearly), as well as summary metadata.
- The "Population data & working notes (xlsx)": includes all of the above, plus interpolated data for every year, plus detailed metadata (with all the information we found so far on everything that happened with each observation from primary observation to final observation). Please note that the file is very large.
- The google spreadsheet for "Total Population": this is merely a copy of the data. There is no additional information to be found here; the spreadsheet exists only to make the data available to the Gapminder World software.
- The google spreadsheet for "Data quality – Total Population": this is merely a copy of the data. There is no additional information to be found here, the spreadsheet exists only make the data available to the Gapminder World software.

Links to these spreadsheets are found at:

http://www.gapminder.org/downloads/documentation/#gd003

Our goal is to have a complete data set, from 1800 to 2008, for "all" countries and territories. Furthermore, we have had the ambition to collect as much meta-data as possible, on an observation-by-observation basis, and to use this information to rate each observation according to its data quality.

The main purpose of the data is to produce graphical presentations in the Gapminder World Graph. Since population is normally used for the size of the bubbles in Gapminder World, a missing observation means that no bubble is shown, irrespectively of whichever other indicator we have in the graph. Therefore, we have included very rough estimates for countries and territories for which reliable data was not available, and we have also, for a few historical observations, simply "guessed".

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.

This is very much a "work in progress". We have tried to organize the work so as to facilitate a continuous improvement of both the data itself and of the documentation. Hence, we would very much welcome comments, suggestions and criticisms.

We have a complete data set for 253 countries and territories, including all the UN-members. In principle, this should include every person in the world. We have complete coverage for the period 1800-2008, and for many countries and territories we have historical data going back even further in time. We have also included projections up to 2030.

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". However, it should be noted that the inclusion or exclusion of any area in this data set does not, in any way, imply a stated opinion of Gapminder as to the legal status of the area.

It is not always clear to what extent certain semi-autonomous or disputed territories are also included in the data for their "mother country", so there may be some "double counting" in these instances. Thus, it is not feasible to simply summarize all the countries and territories in this data set to get the total population of the world since the "double counting" would overestimate the world population.

The rest of this documentation will be organized as follows. First, in section 2, we give an overview of what we have done and how we have rated the data quality of the observations. The following sections, 3 to 7, describe the sources of Gapminder, the primary data and the modifications done to the data. Section 8 summarize the data quality rating and briefly discusses uncertainty ranges. Section 3 to 8 each corresponds to a column in "Population data (xls)".

There are some additional columns in "Population data & working notes (xlsx)" which are described in section 9. Section 10, finally, give some short remarks on the trustworthiness of the meta-data. After that you find table 4, which summarizes all the observations. In the end we have included a reference list, which includes all the sources we indirectly have quoted, to the extent that we are aware of them.

2. What we have done

We started by utilizing existing compilations of international data, where we gave precedent according to how well documented they were (on an observation-by-observation basis) and how good their coverage was. We then turned directly to national sources. Next we turned to "undocumented" sources, such as Wikipedia, the CIA fact book and journalistic accounts. After that we used various extrapolation techniques, and finally after that, we guessed.

For each observation we also tried to include as much meta-data as we could, on an observation-by-observation basis. Most observations have travelled quite a long way from the original primary data, through a long chain of citations and manipulations, where the manipulations by Gapminder, if any, are only the last ones in this chain. We call all those in this chain of citations and modifications "data conveyors".

We have tried to track these chains as far back as we could. This has meant that we had to look up the sources used by our sources, and then the "third layer" of sources used by those

"second layer sources" etc. This work has only started. So far, in only a few cases, have we managed to find the primary source. Many international compilations do not give meta-data or sources on an observation-by-observation basis, only the general principles they've used.

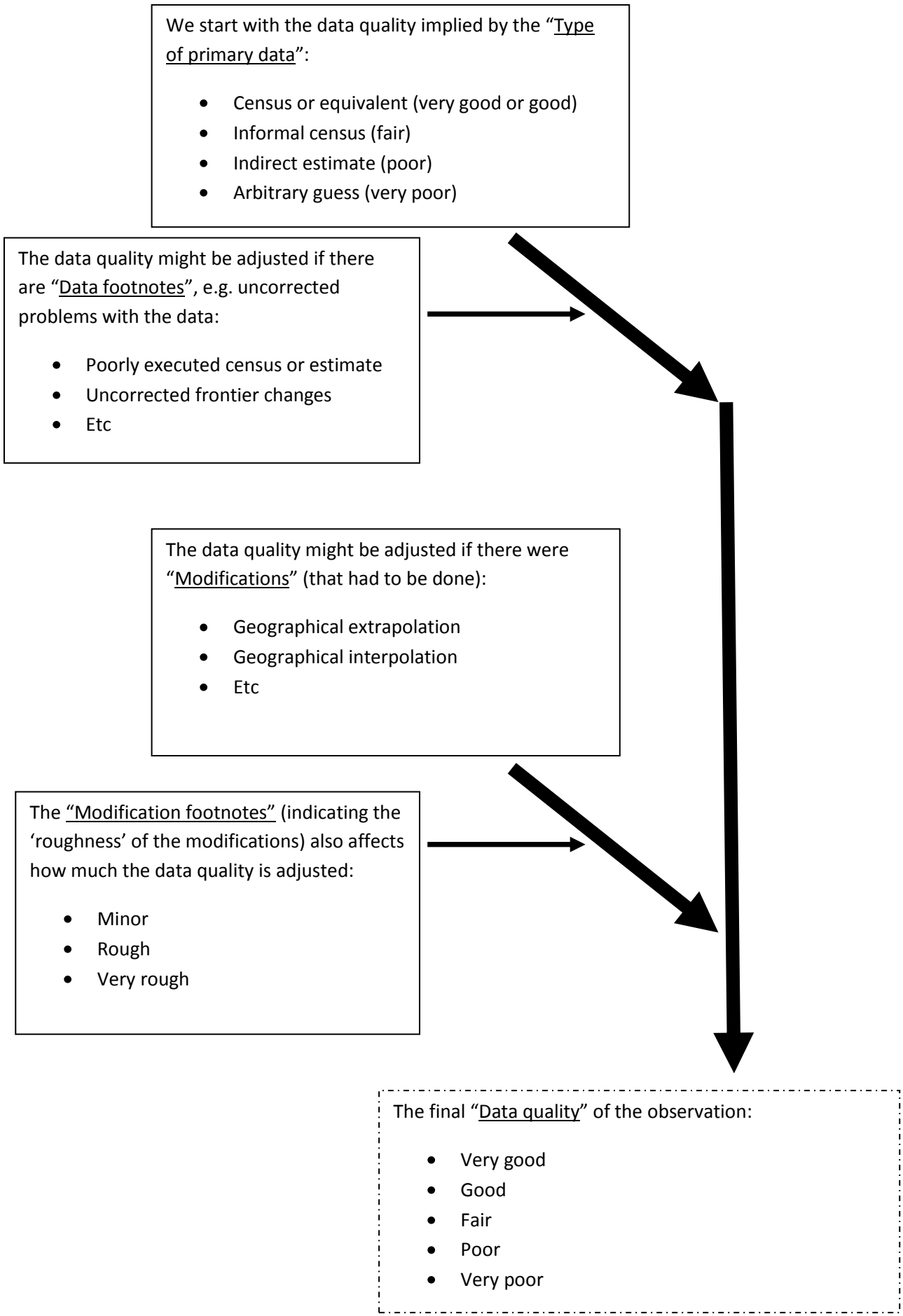
All the meta-data we found, including the chain of data conveyors and the various manipulations done by them along the way, are documented in the excel sheet with working notes. All this meta-data is then summarized in four columns (available in the "population data (xls)"). First we note the "type of primary data", i.e. whether the starting point was a census, some more informal counting of people or an indirect estimate. In the next column we summarize some "data footnotes", e.g. some special merits or (unresolved) problems with the primary data or any other relevant note about the observation.

Third, in the "modifications" column, we summarize all the modifications, if any, done along the way, i.e. both the modifications done by those from whom we took our data and modifications done by us. These have been categorized into a limited set of standard modifications. Fourth, in the column "modification footnotes", we provide further information about the modifications, i.e. whether the modification could be considered minor or rough.

Based on these four types of summary meta-data, we tried to rate the observation into one of five data quality categories: very good, good, fair, poor or very poor. The figure on the next page illustrates how this was done. The "type of primary data" implies a certain data quality, e.g. an observation based on a census is generally of "good" quality while an observation based on an indirect estimate is generally of "poor"-quality.

However, a census can be well or poorly executed and an informal estimate can be made on very wild or very robust assumptions. Things like these will be noted in the "data footnotes", together with other (uncorrected) problems (e.g. unadjusted border changes). If there are such data footnotes of any importance we might adjust the quality rating implied by the type of primary data, e.g. a census might be considered having "fair" data quality rather than "good".

In many cases we or others have tried to correct the problems with primary data by doing some kind of modification, e.g. adjusting for border changes or under-enumeration or interpolation for years without data. These manipulations hopefully improve the quality of the data. However, in only a few cases can we assume that the final data quality is as good as it had been if the problems had not been there in the first place. Hence, a manipulation often implies that the data quality is somewhat lower than it had been if we had been able to use the primary data directly.


We start with the data quality implied by the "Type of primary data":
Census or equivalent (very good or good)
Informal census (fair)
Indirect estimate (poor)
Arbitrary guess (very poor)
The data quality might be adjusted if there are "Data footnotes", e.g. uncorrected problems with the data:
Poorly executed census or estimate
Uncorrected frontier changes
Etc
The data quality might be adjusted if there were "Modifications" (that had to be done):
Geographical extrapolation
Geographical interpolation
Etc
The "Modification footnotes" (indicating the 'roughness' of the modifications) also affects how much the data quality is adjusted:
Minor
Rough
Very rough
The final "Data quality" of the observation:
Very good
Good
Fair
Poor
Very poor

Some modifications require very bold assumptions, which would imply that the data quality rating should be lowered compared to what the type of primary data implies. Other modifications can be considered quite reliable and we can hope to "restore" the quality of the primary data. For example, a census might have excluded nomads, but the nomads were a very small share of the population, and we have a very good sense of what that share is. Hence, the population, after the adjustments for the nomads, could still be considered as having a "good" data quality.

Depending on the type of modifications done (as stated in the column with the same name), we adjust the data quality rating implied by the "type of primary data" and "data footnotes". Some modifications are considered to be more problematic than others. However, a specific modification can also be done in a more or less rough way, which also affects how much we adjust the data quality rating. For a summary of the actual data quality rating see Section 8 and Tables 1 and 2.

Despite this, in the majority of cases we do not have enough meta-data to do this rating, instead they are given the rating "NA", i.e. not available. We have tried to be as systematic and consistent as possible; however, there were still a number of decisions that required us to use our best judgment. We do not claim to have created a "rating blue print" that can be followed mechanically.

We only assign a data quality rating when we have managed to trace both the primary data as well as all the major manipulations done on the data along the way (both by us and by others along the way). One exception to this rule was when the manipulations were so rough that we judged the end result to be so imprecise as to render it the lowest rating, whatever type the original data might have been.

Another exception was the very few cases when we dealt with small islands that were uninhabited during certain years. If we decided that we trusted such a claim we always rated it as "good" since zero is a pretty precise figure.

Note that we will only rate an observation as "very good" when we have found the documentation testifying both to the high quality of primary data collection as well to the absence of problematic modifications. Hence, there is likely to be many observations that will qualify as "very good" when we find more documentation, but are now only rated as being "good".

Before we explain our summary metadata we start by saying something about the sources *we* have used.

3. Sources used by Gapminder

Below we describe the main sources we used. Summary information on this can be found in the "Population data (xls)", in the column "Source of Gapminder".[1]

The sources below are roughly listed in the order of priority we used, i.e. we only turned to the next source in the list when there was no data in any of the previous ones. However, since this dataset has grown in an iterative process, it is not guaranteed that the order of priority has been followed to the letter. Rather, the guiding principle has been a combination of convenience (i.e. coverage), on the one hand, and quality and transparency, on the other.

3.1 The dataset of Angus Maddison

The starting point has been the monumental work of Angus Maddison. His dataset has the best historical coverage (with some series going very far back in history), is available digitally and has a very good written documentation for individual observations. Even though most of the data were available digitally from his homepage, there were also some additional data in his written documentations (e.g. Maddison 2001).

To avoid making our dataset too big, we have not included his earliest observations (e.g. those from the first millennium), but that might be included in future updates.

3.2 Mitchell "International historical statistics"

To fill the gaps in the data from the above source we used Mitchell (1998 a & 1998 b).

Mitchell has, in three publications, compiled historical data for most countries of the world. Mitchell provides two sets of tables for the total population of countries: one which only includes observations based on censuses or similar and one that also includes estimates. We used the first set of tables. We were able to find footnotes on an observation-by-observation basis.

3.3 UN statistics division

To fill the gaps in the data from the above sources, we used data from the UN statistical division. They have yearly data for 1970-2006. We have been unable to find meta-data on an observation-by-observation basis for this compilation.

3.4 World Population Prospects: The 2006 Revision

To fill the gaps in the data from the above sources, we used data from the World Population Prospects (which can also be found through a search on the UN data search engine). They have data for every fifth year from 1950 to 2050. There were some meta-data for some specific observations, but we have been unable to find sufficiently detailed meta-data on an

[1] We include every specific observation in "Population data.xls". In the description in this text we have grouped the sources into categories (although several categories only include one source each).

observation-by-observation basis to reconstruct the full chain from primary data to the final observations.

3.5 U.S. Census Bureau, International Data Base.

To fill the gaps in the data from the above sources, we used data available from the International Data Base at the U.S. Census Bureau. We were unable to find meta-data on an observation-by-observation basis for this compilation.

3.6 "The population of Oceania in the second millennium" by Caldwell et al (2001)

To fill the gaps in the data for the Pacific from the above sources, we used an article of Caldwell et al (2001). We were able to find meta-data on an observation-by-observation basis in this article.

3.7 National sources

To fill the gaps in the data from the above sources, we turned to more direct sources on a country-by-country basis, e.g. the statistical bureaus of individual countries. In many cases this meant that we went directly to a "primary source", e.g. census reports and the like.

3.8 "Undocumented" ad-hoc sources

To fill the gaps in the data from the above sources we turned to a variety of more undocumented sources such as Wikipedia, the CIA fact book and journalistic accounts. We have in general been unable to assess the quality of the data in these sources.

This does not, in itself, imply a low precision, but rather that we are not sure about what the quality is (see the discussion on the "meta-meta-data" below). We have tried to cross-check some of the observations in some of these series with other sources, whenever possible.

3.9 Gap-filling using the sources indirectly

The sources described so far were the only ones used by us. However, there were still some gaps. For these gaps we used various indirect data, based on the sources above, estimated with various interpolation and extrapolation techniques (e.g. what we have labeled geographical interpolation or extrapolation, temporal interpolation etc). See section 6, "Modifications", for more details.

3.10 Arbitrary guesses

When all of resources listed above had been exhausted there still remained a few missing observations to get the full data-set we wanted (especially the desire to have a data-set starting in 1800). For these observations we simply made an arbitrary guess, i.e. setting the observation equal to the earliest observation available. These observations were solely included to enable display of all countries in the graph; they do not add any information what-so-ever.

4. Type of primary data

The "sources" listed above are simply where *we* took our data, i.e. normally one of the major compilations of international data. All that data originate from one primary source or another, and we have tried to track down what that primary data is. We have grouped these primary sources into four categories, which are described below.

4.1 Census or equivalent

These include censuses, administrative enumerations or registration records executed with an elaborate and documented methodology according to modern standards.

If there are no modifications or other problems or merits noted for an observation, then they are rated as having at least a "good" data quality. It is probably the case that many also could be considered "very good", but as long as we have not found the documentation testifying to a very high quality methodology we do not rate them higher than "good".

4.2 Informal census

By informal censuses we mean an actual counting of people, or at least something very similar (e.g. households, taxpayers), but a counting that was not done according to modern standards. This includes pre-modern censuses, where there were some important deviations from modern standards of censuses. This can also include more informal impressions on the number of people living in an area made by contemporary eye-witnesses, but where the conditions were such that the observation could be considered somewhat reliable (e.g. the observation concerns a very small area such as a small island).

If no modifications were done, or no special problems or merits were noted, the observation is rated as being at least "fair".

4.3 Indirect estimate

Indirect estimates do not entail any actual counting of people. Rather, they are based on other indirect information, combined with a more extensive set of assumptions. This could, for example, be based on tax records, the size of armies, archeological evidence of

settlement patterns and size of cities combined with assumptions on how all this could relate to the size of the population. Typically, a number of different kinds of information are combined.

If no modifications are done or special problems or merits noted, then the observation is rated as being at least “poor”.

4.4 Arbitrary guess

When no information was available, however indirect, we arbitrarily set the population as being the same as the earliest observation we had. This is only done to get a full data set for graphical display against other indicators and is only done for the first relevant year (since the Gapminder graph interpolates all other years).

Arbitrary guesses are always rated in the lowest data quality category, i.e. “very poor”.

5. Data footnotes

These notes comment on unresolved problems or merits with the observation. Among the notes of other merits or shortcomings of the data, we can comment on a few. An observation for an “uninhabited” area is always rated as being “good”; if we are reasonably confident that the area really was uninhabited, then zero is pretty exact figure.

Uncorrected frontier changes or definitional changes, i.e. when modifications have not been done to remedy frontier changes or problems with under-enumeration or the like, lowers the data quality rating with one step.

6. Modifications

In many cases the primary data has been modified in one way or another to get the final figure. We have tried to classify the modifications done to the primary data into a limited number of categories. Note that we considered both the modifications we have done, as well as other modifications done by others earlier in the chain of data conveyors.

Some modifications can be considered more problematic than others. Furthermore, each modification can be done in a more or less “rough” *way*. Hence, how the degree of ambiguity of any modification depends on two factors: the type of modification, and how “roughly” the specific modification was done.

We have tried to grade the way each manipulation is done as being either “minor”, “rough” or “very rough”. Note, that based on our grading scale, that, for example, a “rough”

geographical extrapolation can be seen as more problematic than a "rough" geographical interpolation.

What we have said so far is based on a relatively straightforward model: we have some primary data, which someone may modify slightly, and then next along the chain, perhaps someone else may perform some additional modifications on that data, and the result is the figure Gapminder use. In such a case we are as to what we consider our primary data and what the source of our data is.

However, in many cases we utilize more than one observation to get to our final figure. When we do a temporal interpolation, for example, we interpolate between two data points, where each data point has a primary source and which might also have been modified before we used them. Even when we adjust a figure for, say, the non-inclusion of a minority group, this adjustment is based on some information, whether it is our guess or the estimated size of that group in some other year.

Ideally we should consider the data-quality of all our sources, as well as the manipulations done when combining all the information. However, at the present stage of this work, we have, in many cases, only considered the quality of the most important piece of data. The original data that we consider as the most important one we call "source data", and so it is the primary data of the "source data" that is included in our documentation. Exactly what we consider as our source data in cases with multiple pieces of information depends on the type of modification done, and is noted below.[2]

A similar, but easier, problem occurs when manipulations have been done in several steps. Then the observation has been classified according to the "roughest" manipulation (using the principle that a chain is not stronger than its weakest link), e.g. the small European countries for 1820 have first been extrapolated from 12 larger European countries, and then interpolated. These have been labeled "geographical extrapolation, in several steps".

Please find below the categories of modifications:

6.1 Summations of parts

Sometimes we have data for all the constituent parts of a territory, e.g. we have data for Guernsey and Jersey that constitutes "the Channel Islands". Then it is simply a matter of adding up these observations to get the new one. This should not cause any major problems. If the constituent parts have a uniform quality and there are no other problems noted, then the modification is considered "minor" and the resulting observation should have the same quality as the original observations.

[2] In future versions we hope to deal with this issue in a more consistent and smooth way.

6.2 Larger area minus non-included parts

Sometimes we have the total population for a larger area, as well as the population of some of the constituent parts, but we lack data for the other constituent parts. For example, we might have data for "Serbia and Montenegro" and for "Serbia", but no data for "Montenegro". Then we could simply calculate the population of Montenegro as "Serbia and Montenegro" minus "Serbia".

Ideally this should not pose any problems, just as for "summation of parts". However, inconsistencies between the observations can pose a much more serious problem than for "summations of parts", especially if the area we are looking for is a very small part of the larger area we do have data for. For example, the population of Timor Leste is less than 1% of Indonesia, which is actually smaller than the deviations between some of the sources for Indonesia. Hence, if the population of Timor Leste were estimated as "Indonesia including Timor Leste" minus "Indonesia excluding Timor Leste" (it is not), and the two population figures came from different sources, then we might, in the worst case, even end up with a negative population for Timor Leste.

Accordingly, such a modification could lower the quality of the data. The extent to which we consider this manipulation to be "rough" depends on a number of factors, including:

- Whether or not the population of our area of interest constitutes a major share of the population of the larger area.

6.3 Geographical interpolation

With "geographical interpolation" we assume that the country's population has had the same population growth as a larger area of which it is a part. This could be used when we have data for former countries that have now split up, while at the same time, we have data for the new countries, but only for a few years. Some of the sources (e.g. Maddison & Caldwell) supply more aggregated data, e.g. total population for a whole region or for a group of countries.

If we assume that the share of a specific country in the larger area was the same in earlier years then we can use the regional total to estimate the population of the country in the earlier years for which only regional data is available. As an example:

$$Population\ of\ Slovenia\ (1910) = \\ = Population\ of\ Yugoslavia\ (1910) \\ \times \frac{Population\ of\ Slovenia\ (1950)}{Population\ of\ Yugoslavia\ (1950)}$$

In the above case we consider the population of Yugoslavia (1910) as being the "source observation".[3] Hence, the information about "Type of data" and the like would, in this case, refer to the population of Yugoslavia in 1910.

The term "interpolation" (which we use in a rather new way) refers to the fact that our country of interest is a constituent part of the larger area we are using. Therefore, we at least know that the population of our country is less than the population of the larger area (assuming that the population of the larger area is correct, of course). This means that geographical interpolation is a somewhat better method than the geographical extrapolation discussed below.

The extent to which we consider this manipulation to be "rough" depends on a number offactors, such as:

- The number of years for which we assume that the population share is constant.
- Whether or not the country's population constitutes a major share of the larger area.
- Whether or not we have some general historical accounts indicating that there were large population movements.

If the interpolation is only minor (e.g. short time period, our country is not a very small part of the larger area and no indications of large population movements) then the observation might retain the data quality rating implied by the type of the primary source.

If the interpolation could be considered as "rough" (and we have no indications of other problems) we lower the data quality rating one step.

If the interpolation could be considered "very rough", or as "rough" with some other problematic recalculations, then we lower the data quality rating two or more steps, depending on the nature of those other recalculations.

6.4 Geographical extrapolation

When doing a "geographical extrapolation" we assume that the country's population has had the same population growth as a neighboring country. In principle we mean the same thing as with "geographical interpolation", the only difference being that our country is not a part of the area we used for our estimation. As an example:

$$Population\ of\ Estonia\ (1820) = \\ = Population\ of\ Lithuania\ (1820) \times \frac{Population\ of Estonia\ (1860)}{Population\ of\ Lithuania\ (1860)}$$

[3] Since that is an observation for the same year as our observation of interest.

In the above case we consider the population of Lithuania (1820) as being the "source observation"[4]. Hence, the information about "Type of data" and the like would, in this case, refer to the population of Lithuania in 1820.

The term "extrapolation" refers to the fact that we are using data from an area that is not overlapping our country of interest. This means that even if the population data for our "source country" is absolutely accurate, we do not have a "guaranteed maximum" population. This would only be a significant problem if one of the countries has gone through some major migration movements or wars, and the other has not. We tried to minimize this risk by doing some quick review of the country's history. One case in point is 19th century southern Africa where there has been a great deal of population movement; for the time being, this is something for which we are occurred for which we are lacking even tentative information.

The degree of "roughness" for this manipulation depends on a number of factors, including:

- The number of years for which we assume that the population ratio is constant.
- The extent to which we can assume that the two countries had similar population growth rates, e.g. whether we have some general historical accounts indicating that there were large population movements or not.

If this manipulation has been judged to be "very rough", then the observation is always rated as being "very poor", whatever the quality of the "source data" is, since the link between the source data and the observation is based on very speculative assumptions.

6.5 Temporal interpolation

"Temporal interpolation" is the term we use for what is normally meant by interpolation, i.e. drawing a straight line between two points. This can be done in a variety of ways, e.g. either assuming a constant growth rate between two years with data, or simply assuming that the population changed with a fixed absolute number.

Both observations used for the interpolation are considered as the "source data" in this case.

We have not done any temporal-interpolations, with some very few exceptions, since the graphing software does temporal interpolations automatically. However, some of the data conveyors have done temporal interpolations. Furthermore, in the Excel file called "Population data & working notes (xlsx)" we have included observations for all years, with the missing years filled with interpolated values using constant growth.

The degree of "roughness" for this manipulation depends on a number of factors, such as:

- The number of years that have been interpolated

[4] Since that is an observation for the same year as our observation of interest.

- Whether we have some historical accounts that (informally) indicate major upheavals in the population growth, i.e. unique happenings that do not fit into a smooth population movement

If this manipulation has been judged to be “rough”, we lower the data quality rating one step.

6.6 Temporal extrapolation

By “temporal extrapolation“ we mean all other methods of extending a series outside its time range. This can be done in a number of ways; the roughest is to just assume that the population is the same as the closest observation (similar to our “arbitrary guess”). A slightly more sophisticated method is to extend the growth rate of some adjacent period backward or forward. There are also more sophisticated modeled projections, based on various assumptions of fertility and the like.

Generally speaking, this manipulation is rougher than temporal interpolation since we only have either a starting or ending year, while in temporal interpolation we have both.

The observation(s) considered as the “source data” are all the observations used as a basis for the extrapolation. Hence, if we simply have assumed that the population is the same as for some other year, then this other year is considered the sources data. On the other hand, if the constant growth rate over a certain period has been used, then all the observations used to calculate the growth rate are considered the source data.

In principle, Gapminder has not done any temporal extrapolations. The few exceptions are cases for some very short time spans, when we were not sure about what year the data were referring to. Furthermore, “arbitrary guess” could, as noted, be considered as being a very rough form of extrapolation.

However, the other data conveyors have occasionally done temporal extrapolations. We have not always managed to find information on exactly what observations are based on this method, and exactly how that was done. Observations for future years can of course be assumed to be based on this method.

The degree of “roughness” for this manipulation depends on a number of factors, such as:

- The number of years that have been extrapolated
- The sophistication of the methods used
- Whether we have some historical accounts that (informally) indicate major upheavals of the population growth, i.e. unique happenings that are difficult to capture in models

If this manipulation has been judged to be “minor”, we lower the data quality rating one step. If this manipulation has been judged to be “rough”, we lower the data quality rating

two steps. If this manipulation has been judged to be “very rough”, then the observation is always rated as being “very poor”.

6.7 Adjustments for under-enumeration

Sometimes the data fail to cover certain groups, e.g. nomads, or have been judged, for various reasons, that a census had less than ideal coverage. In those cases efforts have been made to remedy these shortcomings by adjusting the figures. Sometimes, when specific groups have been missed (such as our friends, the nomads) attempts have been made to estimate the size of these groups, e.g. by doing some kind of interpolation or utilizing other data. Other times more ad-hoc adjustments have been made.

Note that when the excluded groups are geographically based, we consider any adjustments to be “recalculations to fit present borders”, which is described below.

The degree of “roughness” for this manipulation depends on a number of factors, such as:

- The estimated relative size of the excluded group
- The method of estimating the size of the excluded group. Some method should be almost free of problems, e.g. if we happen to have exact numbers from some other source (in which case it is a simple matter of adding two numbers together), other could be considered more rough, e.g. some indirect estimate or interpolation, while others could be based on some arbitrary adjustments

If this manipulation has been judged to be “rough”, we lower the data quality rating one step.

6.8 Recalculated to fit present borders

As can be noted in the document “Countries & Territories in Gapminder World” we try to make the historical data refer to the present borders of a territory, even when there have been substantial border changes in the past. Hence, there is often a need to do some recalculations, e.g. to subtract the population in areas that are no longer part of the country, and to add population for areas that were not part of the country in the past.

Ideally we have the data for the relevant sub-regions (that constitute the difference from the present borders). In such cases it is a simple matter of adding or subtracting the sub-regions in question. This should ideally not have any major impact on the uncertainty range rating for the resulting observations.

However, in most cases we do not have direct data for some of the relevant areas, so we have to get the needed data in more indirect ways by utilizing some of the of the methods described above, e.g. geographical interpolation.

The degree of “roughness” for this manipulation depends on a number of factors, such as:

- The relative size of the sub-areas that have to be estimated indirectly.
- The method of estimating the various sub-areas. Judging the roughness of these estimation methods has already been discussed under each method.

If this manipulation has been judged to be “minor”, it does not affect the uncertainty range rating except that it can, at most, be rated as being “good”. If this manipulation has been judged to be “rough”, we lower the data quality rating one step.

7. Modification footnotes

The modification footnotes are of various kinds and are more ad-hoc in nature. Hence, we will not give the full list of these footnotes here (such a list can, indirectly, be found in table 4 below). The notes comment on the “roughness” of the modifications done, if any.

All the notes on the roughness of a manipulation has the same structure: it begins with either “minor”, “rough” or “very rough” followed by the motivation for the roughness rating within a parenthesis.

One major factor determining the roughness of the interpolations and extrapolations has been the number of years for which an assumed growth rate has been applied. For example, if we assume that the share of Estonia’s population in the total of USSR was the same in 1800 as in 1975, a period of 175 years, this is obviously quite a rough assumption.

As a rule of thumb we have rated extrapolations and interpolations as “minor” if the time period in question was 10 years or less, as “rough” if it was 50 years or less (but more than 10 years), and as “very rough” if the time period was more than 50 years.

Another factor that yielded a rougher grading for the modification was if there was a series of modifications. For example, sometimes there was first a geographical extrapolation of a group of countries, and that figure was used to make geographical interpolation to an individual country. These cases have been noted with “… in several steps”.

8. The data quality rating

As stated earlier we have put each observation into one of five data quality ratings: very good, good, fair, poor, very poor. To display these in the graph we have coded them with a number, with “5” representing “very poor” and “1” representing “very good”.

The principles for assessing the data quality of each observation are summarized in Tables 1 to 3 below. Note, however, that the rating has not been done in a mechanical way. In many cases ad-hoc judgments have been made, and the principles outlined in Tables 1 to 3 have not always been followed exactly: Rather, the principles have been used as general

guidelines. More specifically, in several instances it was felt that a range of problems did not "add-up" on each other, e.g. once an observation had been judged "poor" one additional modification is not sufficient to render it being "very poor".

Also note that all possible combinations have not been filled in yet, as the need for some of them has not yet arisen and this still is a work in progress. For descriptive statistics of the number of observations in each category, see Table 4 below.

Type of primary data	**"Initial" data quality**
Census or equivalent	"good" or "very good"
Informal census	"fair"
Indirect estimate	*"poor"*
Arbitrary guess	*always "very poor"*

Table 1. Data quality implied by the type of primary data

Data footnotes	**Effect on data quality**
Uninhabited	always "good"
Uncorrected frontier changes or definitional changes	lower one step

Table 2. Some examples of the effects of "data footnotes" on the data quality

		Modification footnotes (the roughness of the modifications)		
		minor	Rough	very rough
Modifications	Summations of parts	none		
	Larger area minus non-included parts			
	Geographical interpolation	none	lower one step	lower two steps or more
	Geographical extrapolation			always “very poor”
	Temporal interpolation		lower one step	
	Temporal extrapolation	lower one step	lower two steps	always “very poor”
	Adjustments for under-enumeration		lower one step	
	Recalculated to fit present borders	"good" is maximum	lower one step	

Table 3. The effects of modifications on the data quality

Our long term ambition is to eventually transfer these data quality ratings into uncertainty ranges, i.e. to provide a likely maximum and minimum value. There still remains much work to do on this, but we could state some general principles. Just to get a starting point, however, we could cite Mitchell (1998 a, p. 1):

“It is generally agreed by demographers that the almost universal tendency of censuses is to under-enumerate, and the probability is that this was more pronounced in earlier than in later censuses. … Perhaps it can be regarded as reasonable to assume that regular censuses, other than the first one or two of a series, are accurate to within less than ten percent overall, and probably a good deal less than that in many cases, especially in developed countries. Isolated censuses may well have higher margins of error, most especially in large countries.”

We envision that our “good” observations roughly correspond to the “regular censuses” and “very good” roughly corresponds to those that have “good deal less” errors.

“Fair” observations have very large margins of errors but give a fairly reliable indication of the order of magnitude, e.g. margins of error on the scale of 50%.

“Bad” only gives a rough order of magnitude, i.e. are we speaking of thousands or hundreds of thousands.

“Very bad” observation does not in principle contain any information beyond what everyone already knew, e.g. that Malawi did not have billions of inhabitants in 1800.

9. Other information that is only available in “Population data & working notes (xlsx)”

The Excel sheet labeled “Population data & working notes (xlsx)” contains all the columns included in “Population data (xls)”. In addition, it also includes a number of other columns with supplementary information. These columns are arranged so that the columns to the right, in principle, are one step further back in the chain of data conveyors.

Thus, first (starting from the left on the sheet) are a couple of columns describing what modifications Gapminder has done with the data, as well as other notes made by us. Then comes “source by observation” which are the sources used by us. Then come some general notes on geographical coverage. Then come the “sources of our sources”, which we call “second layer of sources”. Finally we have the “sources of the sources of our sources”, which we call the “tertiary layer of sources” (any suggestions on a more smooth terminology in this area is more than welcomed).

For technical reasons we had to divide long strings of texts into several cells, which meant that for some of the information we have several columns. Hence, when the column title ends with (col 1), (col 2) or (col 3) the bordering cells in these columns are meant to be one string of text.

Here is the full list of the columns in the “Population data & working notes (xlsx)” that are not found in “Population data (xls)”:

- Population – with interpolated values
- Gapminder modifications
- Gapminder modifications - explanations
- Gapminder modifications - explanations (col. 2)
- Other notes by gapminder (col 1).
- Other notes by gapminder (col 2).
- Source of Gapminder
  (NOTE – this column is also included in “population data (xls)” as well)
- source of Gapminder - details (col 1)
- source of Gapminder - details (col 2)
- Geographical coverage (col 1)
- Geographical coverage (col 2)
- Secondary layer of sources (col 1)
- Secondary layer of sources (col 2)
- Tertiary layer of sources (col 1)
- Tertiary layer of sources (col 2)
- Tertiary layer of sources (col 3)

10. About “meta-meta data”

Since we depends on the works of others along the data pathway in almost all cases, we are not always certain that we have understood exactly what the raw data was, and exactly what modifications have been done. Therefore, the quality of the meta-data might also vary from observation to observation. Accordingly, we could also include some kind of meta-meta data, but we have chosen not to do so at this stage.

Type of primary data	Data footnote	Modifications	Modification footnote	Very good	Good	Fair	Poor	Very poor	NA
Census or equivalent	None	None	None		275				
		Recalculated to fit present borders	Minor (recalculations include geographical extrapolation, but only for minor parts of country and within 10 years)		1				
			Minor (recalculations include geographical interpolation, but only for minor parts of country and within 10 years)		1				
		Summations of parts	None		1				
		Adjustments for underenumeration	Rough (adjustment represent more than 5% but less than 20% of total)			2			
		Temporal extrapolation	Minor (10 years or less)			1			
	Census had only some minor shortcomings.	Geographical interpolation	Rough (more than 10 years, 50 years or less, other adjustments also done)		2				
	Census had some shortcomings.	Geographical interpolation	Rough (more than 10 years, 50 years or less, other adjustments also done)			2			
	Some uncorrected definitional changes	None	None			1			
	Some uncorrected frontier changes	None	None			11			
		Recalculated to fit present borders	Minor (recalculations based on direct data)			1			
			Minor (recalculations include geographical interpolation, but only for minor parts of country and within 10 years)			2			
	Census extrapolated with register data of vital events.	Geographical interpolation	Rough (less than 10 years, major population movements in between, other adjustments also done)			2			
	Census extrapolated with post 1922 growth rate and register data of migration.	Temporal extrapolation	Rough (less than 10 years, turbulent years, geographical interpolation and other adjustments also done)				2		

Type of primary data	Data footnote	Modifications	Modification footnote	Very good	Good	Fair	Poor	Very poor	NA
Informal census	None	None	None			10			
		Geographical interpolation	Rough (more than 10 years, 50 years or less, proportions based on borders that does not match present ones perfectly, other adjustments also done)			76			
		Temporal extrapolation	Rough (more than 10 years, 50 years or less, geographical interpolation and other adjustments also done)				4		
			Very rough (more than 50 years or less, geographical interpolation and other adjustments also done)					2	
	Contemporary eye-witness	None	None			1			
	Guess based on historical accounts	None	None				1		
	Raw data based on variety of sources, both informal censuses and indirect estimates	None	None				6		
	Uninhabited	None	None		2				

Type of primary data	Data footnote	Modifications	Modification footnote	Very good	Good	Fair	Poor	Very poor	NA
Indirect estimate	None	None	None				1		
		Geographical interpolation	Rough (more than 10 years, 50 years or less)				6		
			Rough (more than 10 years, 50 years or less, in several steps)				4		
			Very rough (more than 50 years)					6	
			Very rough (more than 50 years, in several steps)					2	
	Historical account based on variety of sources	None	None				1		
	Numbers disputed	None	None					2	
Arbitrary guess	None	None	None					174	

Type of primary data	Data footnote	Modifications	Modification footnote	Very good	Good	Fair	Poor	Very poor	NA
NA	NA	NA	None						18032
		Geographical extrapolation	Minor (10 years or less)						11
			Rough (more than 10 years, 50 years or less)						44
			Very rough (more than 50 years and distant country)					2	
			Very rough (more than 50 years)					3	
			Very rough (more than 50 years, in several steps)					20	
		Geographical interpolation	Minor (10 years or less)						21
			Rough (less than 10 years, but combined with other recalculations)						12
			Rough (more than 10 years, 50 years or less)						48
			Rough (more than 10 years, 50 years or less, combined with other recalculations)						30
			Rough (more than 10 years, 50 years or less, in several steps)						2
			Very rough (more than 50 years)						78
			Very rough (more than 50 years, combined with other recalculations)						1
			Very rough (more than 50 years, involving many countries and some arbitrary assumptions)					37	
		Larger area minus non-included parts	Rough (the population of the larger area estimated)						1
			None						16
		Recalculated to fit present borders	Minor (recalculations include geographical interpolation, for sizeble part of country, less than 10 years)						23
			Rough (recalculations include geographical interpolation, for sizeble part of country, more than 10 years, 50 years or less)						78
			Very rough (recalculations include geographical interpolation, for sizeble part of country, 50 years or more)						106
		Summation of parts	None						123
		Temporal extrapolation	Minor (10 years or less)						15
			None						473
		Temporal interpolation	None						342

Type of primary data	Data footnote	Modifications	Modification footnote	Very good	Good	Fair	Poor	Very poor	NA
NA (continued)	Some uncorrected frontier changes	NA	None						318
		Temporal extrapolation	None						2
	Source gives approximate figures only	NA	None						3
	This observation is probably based on a more limited definition than for 2001 and earlier	NA	None						5
	Uncorrected frontier changes	Geographical interpolation	Rough (more than 10 years, 50 years or less)						3
	Uninhabited	NA	None		4				

Table 4. The number of observations, tabulated by the summary metadata.

References

"Visit and learn" (X), Downloaded 2008-06-02 from: www.visitandlearn.co.uk/LocationFactfiles/TheFalklandIslands/History/tabid/206/Default.aspx

ABC.net (X), Downloaded 2008-04-28 from: www.abc.net.au/nature/island/ep2/about4.htm

Administration of Norfolk Island (X), Census of Population and Housing

Albornoz (1986), "The population of Latin America, 1850-1913" in Bethell (ed)., "The Cambridge history of Latin America", vol. 4, p. 122.

Altman (1988), "Economic growth in Canada, 1695-1739: Estimates and analysis", William and Mary quarterly.

Annuaire statistique de la France (1966).

Annual movement from Ritzmann-Blickenstorfer (1996), Historical statistics of Switzerland

Annuiaire statistique de la Belgique et du Cono Belge (1955)

Australian Government, Bureau of Metereology. Downloaded 2008-06-02 from www.bom.gov.au/weather/wa/cocos_island/history.shtml

Aye Hlaing (1964), "Trends of economic growth and income distribution in Burma 1870-1940", Journal of the Burma research society.

Baptista (1991), "Bases de la Economia Venezolana 1830-1989"

Bardet & Dupaquier (1997), "Histoire des populations de l'Europe"

Befolkningsstatistikk, Statistisk sentralbyrå & Sysselmannen på Svalbard & Polish Academy of Sciences.

Beloch (1961), "Bevölkerungsgeschiste Italiens", pp. 351-4.

Bethel (ed.) (1985), "The Cambridge history of Latin America", vol. III

Bethell, ed. (1986), "The Cambridge history of Latin America", vol. IV

Brundenius (1984), "Revolutionary Cuba: the challenge of economic growth with equity"

Bulmer-Thomas (1987), "The political economy fo central america since 1920"

Butlin (1998), "Our 200 years".

Caldwell et al (2001), "The population of Oceaneia in the second millenium", downloaded 080606 from http://htc.anu.edu.au/pdfs/Oceania%20manuscript.pdf

Carreras (ed.) (1989), "Estadisticas Historicas de Espana: Siglos XIX-XX".

CBS (1970), "Zeventig jaren statistiek in tijdreeksen".

CBS, Oslo (1969), "Historical Statstics, 1968".

CEPAL (ECLAC), statistics division (1962), "Cuadros del producto interno bruto a precios del mercado en dollares de 1950", Santiago (mimeo)

Christmas Island Tourism Association. Downloaded 2008-04-28 from: www.christmas.net.au/history.php

CIA World Factbook. Downloaded 2008-04-28 from: www.cia.gov/library/publications/the-world-factbook/fields/2119.html.

Cook (1981), "Demographic collapse: Indian Peru, 1520-1620"

de Vries (1984), "European urbanization 1500-1800"

Deane & Cole (1964), "British economic growth 1688-1958"

Dennoon (1997), "New Economic Order: Land labour and dependency", in Denoon (ed.), "The Cambridge History of the Pacific Islanders"

Dickson, Grada & Daultry (1982), "Hearth tax, household size, and Irish population change 1672-1821" in "Proceedings of the Royal Irish Academy", vol. 82, C, no 6, p. 156.

Dickson, O Grada & Daultrey (1982), "Hearth tax, household size and Irish population change 1672-1981", proceedings of the Royal Irish Academy

Eisner (1961), "Jamaica, 1830-1930: A study in Economic Growth".

Engerman & Higman (1997), "The demographic situation of the carribean slave societies in the eighteenth and nineteenth centuries", in Knights (ed), "General History of the Carribean", vol III, UNESCO, pp. 50-7

Eric Solsten (ed.) (1991), “Cyprus, a country study”, Library of Congress, Appendix A, table 5. Downloaded 2008-06-10 from http://lcweb2.loc.gov/frd/cs/cyprus/cy_appen.html

ESCAP (1983), "Migration, Urbanization and Development in South Pacific countries".

Farell (1991), "History of the Northern Mariana Islands".

Federal statistical office, Bern (1953), "Annuaire statistique de la Suisse 1952"

Feinstein (1972), "National Income Expenditure and Output of the United Kingdom 1855-1965"

Greer (1966), "The demographic impact of the Mexican revolution 1910-21", Master thesis, University of Texas

Hansen (1974), "Ökonomisk vaekst i Danmark", vol II, Institute of Economic history, Copenhagen

Hawke (1985), "The making of New Zeeland", pp. 10-11

Henry & Blayo (1975), "La population de la France de 1740 à 1860", in Population, november 1975, pp 97-9

Higman (1984), "Slave populations of the Brittish carribean, 1807-1834", p. 417.

Hofman (1992), "International estimates of capital: a 1950-1989 comparison of Latin America and the USA", research memorandum 509, Institute of Economic research

Huntsman & Hooper (1996), "Tokelau: A historical ethnography".

IBGE (1960), "O Brazil em Numeros", p.5

Index Mundi. Dowloaded 2008-04-28 from: www.indexmundi.com

INE (X), "Espana: annuario Estadistico 1977", p. 49.

INEGI (1985), "Estadísticas Históricas de México", vol I, p. 311

Istat, Rome (1976), "Sommario di statistiche storiche del Italia", 1861-1975.

Kausel, (1985), "150 jahre Wirtschaftswachstum"

Kausel, A., "Österreichs Volkseinkommen 1830 bis 1913" in Gschichte und Ergebnisse der zentralen amtlichen Statistik in Österreich 1829-1979, Beitrage zur österreichischen Statistik, Heft 550, Vienne (1979), pp. 692-3

Kirsten, Buchholtz & Köllmann (1956), "Raum und Bevölkerung in der Weltgeschichte"

Krantz (1988), "New estimates of wedish historical GDP since the beginning of the nineteenth century", Review of income and wealth, june 1988

Kwon & Shin (1977), "On population etimates of the Yi dynasty, 1392-1910", Tong-a Munhwa

Lee, "On the accuracy of pre-famine Irish censuses" in Goldstrom & Clarkson (1981), Irish population, economy and society, p. 54.

Leff (1982), "Underdevelopment and development in Brazil", vol. 1

Lorimer (1946), "The population of the Soviet Union: history and prospects"

Luis Bertola and associates (1998), "PBI de Uruguay 1870-1936" plus supplementary information supplied by Luis Bertola

Maddison (1971), "Class structure and economic growth", pp. 164-5.

Maddison (1989), "Dutch income in and from Indonesia, 1700-1938", Modern Asian Studies, pp. 645-70

Maddison (1991), "Dynamic forces in capitalist development"

Maddison (1998), "Chinese Economic performance in the long run"

Maddison (2001), "The world economy - a millenial perspective"

Maddison (2003), "The World Economy - historical statistics"

Maddison on-line. Downloaded 2008-03-14 from: www.ggdc.net/maddison/ (The file last updated: August 2007).

McArthur (1967), "Island Populations of the Pacific"

McCarthy (1990), "The population of Palestine"

McEvedy & Jones (1978), "Atlas of world population history"

Mead (1967), "Growth and structural change in the Egyptian economy", pp. 295, 302.

Metz, Helen Chapin (ed.) (1994), “Mauritius”, Library of Congress, the chapter on “early settlement”. Downloaded 2008-06-12 from http://lcweb2.loc.gov/cgi-bin/query/r?frd/cstdy:@field(DOCID+mu0012)

Mitchell (1962), "Abstract of British Historical Statistics"

Mitchell (1975), "European Historical statistics 1750-1979"

Mitchell (1975), "European Historical statistics 1750-1979"

Mitchell (1983), "International historical statistics: the Americas and Australia"

Mitchell (1998 a), "International historical statistics - Africa, Asia & Oceania 1750-1993", 3rd edition.

Mitchell (1998 b), "International historical statistics - The Americas 1750-1993", 4th edition.

National population and family planning commission of China (X), "?". Downloaded 2008-06-04 from: www.chinapop.gov.cn/rkzh/rk/sjrkdt/t20040326_46901.htm

Nelson & Nelson (1992), "The island of Guam".

Norfolk Island, Census 1981

Norfolk Island, Census 1986

Norfolk Island, Census 1991

Norfolk Island, Census 1996

Norfolk Island, Census 2001

Nunes, Mata, Valerio (1989), "Portuguese Economic Growth, 1833-1985", Journal of European Economic History, fall

Nutter (1962), "The growth of the industrial production in the Soviet Union", NBER

O Grada in Bardet & Dupaqiuier (1997), "Histoire des populations de l'Europe", vol 1, p. 386.

Office of the Governor of Svalbard (?) (2008), "Historical Background", downloaded 080604 from: www.sysselmannen.no/hovedEnkel.aspx?m=4529

Oliver (1989), "Oceania: The Native cultures of Australia and the Pacific Islands", Vol 1.

Prados (2002), "El progreso Economico de Espana, 1850-2000"

Public Record Office, England Colonial Office Correspondence/Straits Settlements & John Hunt: 'Reaction to Mistreatment of Chinese Coolie on Christmas Island 1900-04' Litt, B thesis, ANU 1982

Public Record Office, England War Office and Colonial Office Correspondence/Straits Settlements & J. Pettigrew: 'Christmas Island in World War II ' Australian Territories January 1962

Raychaudhuri & Habib (1982), "The Cambridge economic history of India"

Republic of Cyprus, Ministry of Finance (1989), “Statistical Abstract, 1987 and 1988.

Rogers (1995), "Destiny's landfall: A history of Guam".

Rosenblat (1945), "La poblacion Indigena de America Desde 1420 Hasta la Actualidad"

Rosenzweig. Ed. (no date, 1960?), "Fuerza de travajo y actividad economica por sectores, estadistícas dell Porfiriato"

Rusell (1998), "Tiempon I Manmofo'na: ancient chamorro culture and history of the Northern Mariana Islands"

Sanchez-Albornoz (1986), "Cambridge hitory of Latin America, vol. IV"

Shepherd & Beckles, eds (2000), "Carribean slavery in the Atlantic world"

Sivasubramonian (2000) "The national income of India in the twentieth century"

Smith (1983), "Niue: the Isand and its people".

Smith, L. (1980), "The Aboriginal Population of Australia"

Smits, Horlings, van Zanden (2000), "Dutch GDP and its components, 1800-1913".

Sompop Manarungsan (1989), "Economic development of Thailand, 1850-1950", PhD thesis, University of Groningen

St. John-Jones (1983), “The Population of Cyprus”, 33

Statistics Norway, "NOS D 330 Svalbardstatistikk 2005", Table 45: Personer i bosetningene 1. januar. 1990-2005. Dowloaded 2008-04-28 from: www.ssb.no/emner/00/00/20/nos_svalbard/nos_d330/tab/045.html

Suh (1978), "Growth and structural change in the Korean economy, 1910-1940"

Svennilson (1954), "Growth and stagnation in the European Economy"

Taeuber (1958), "The population of Japan"

Thornton (1987), "American indian holocaust and survival: a population history since 1492"

Turpeinen (1973), "Ikaryhmittainen kuolleisus Suomessa vv. 1751-1970".

U.S. Census Bureau, International Data Base. Downloaded 2008-03-14 from: http://www.census.gov/ipc/www/idb/tables.html

Ubelaker (1976), "Prehistorics new world population size: historical review and current appraisal of North American estimates", American Journal of physical anthropology

UN (1952), "Demographic yearbook 1951".

UN (1960), "Demographic yearbook 1960".

UN (1960), "Demographic yearbook 1960".

UN (X), "Demographic Yearbook"

UN Population Division, World Population Prospects, 2006

UN statistics division. Downloaded 2008-03-14 from: http://unstats.un.org/unsd/snaama/selectionbasicFast.asp

United Nations General Assembly (UN Document Symbol: A/AC.109/2002/2)

US bureau of the census (october 2002), International Programs Center

US department of commerce (1975), "Historical statistics of the United States, colonial times to 1970"

Valerio (2001), "Portuguese Historical Statistics", vol 1

Wikipedia. "Saint Barthélemy". Downloaded 2008-04-28 from: en.wikipedia.org/wiki/Saint_Barth%C3%A9lemy#cite_note-population-1.

Wikipedia. "Cocos (Keeling) Islands" Downloaded 2008-06-02 from: http://en.wikipedia.org/wiki/Cocos_(Keeling)_Islands

World Population Prospects: The 2006 Revision. Downloaded 2008-03-14 from: http://data.un.org/Data.aspx?q=population&d=PopDiv&f=variableID:12 .

Wrigley, Davies, Oeppen, Schofield (1997), "English population history from family reconstitution 1580-1837"

X (X) "Bermuda's History" at the homepage www.bermuda-online.org , downloaded from homepagae 2008-06-02.

X (X), "Cambridge history of Latin America, vol III".

X (X), "Japan statistical yearbook 1975"

X (X), "Norfolk Island … be surprised", downloaded 2008-06-02 from http://www.norfolkisland.com.au/history_and_culture/time_line.cfm .

X (X), Dowloaded 2008-06-02 from: www.government.pn/Pitcairnshistory.htm

Yeatts (1943), "Census of India 1941, vol 1, part I.

Zambardio (1980), "Mexico's population in the sixteenth century: demographic anamoly or mathematical illusion?", Journal of interdiciplary history.

Ålands statistik och utredningsbyrå. Downloaded 2008-04-28 from: www.asub.ax/archive.con?iPage=12&art_id=499

ÅSUB (Ålands statistik- och utredningsbyrå), Befolkningsregistercentralen.