

A COLLECTION OF TANZANIA REGIONAL DATA

INDEX AND EXPLANATORY NOTES

The data listed below was assembled and processed by Hugh Wenban-Smith, with the assistance of Angela Ambroz and Anne Laski (IGC in-country economists) and Gloria Mapunda (Research Assistant), for the purposes of the IGC project “Population growth, internal migration and urbanisation in Tanzania, 1967-2012”. The full reports on this project can be accessed on the [IGC website](#). To obtain a copy of one or more of these data sets, please send your request to the author (wenban@globalnet.co.uk). Use of this data may be cited as Wenban-Smith, HB (2015) *A Collection of Tanzania Regional Data*, together with the index number. The author would be pleased to be notified of any errors users may detect.

I. INDEX

A. Census Related

- A1. The regional structure of mainland Tanzania, 1967-2012
- A2. TzTownPops67-12 (3 sheets)
- A3. TzUrbPops2012 (2 sheets)
- A4. Tanz UrbanPops 1967-2002 (4 sheets)
- A5. Tz Urb&Mig Propensities 78-88, 88-02 & 02-12 (2 sheets)
- A6. Dar_wards_pops_areas_2002 (2 sheets)
- A7. Dodoma_wards_pops_areas_2002&2012 (4 sheets)
- A8. Dodoma wards Shapefiles (7files, zipped)
- A9. 2012 Census Regions_Wards PopsRur&Urb (30 files, zipped – one for each region, incl. Zanzibar)

B. Other

- B1. Crop areas 81, 88, 02-3, 06-7 (2 sheets)
- B2. Road distances (2 sheets)
- B3. Palmer Drought Severity Index Tz (4 sheets)
- B4. Tz Employees & Earnings 67,78, 88, 02 & 12 (7 sheets)

II. EXPLANATORY NOTES

A. Census Related

A1. The regional structure of mainland Tanzania

This note documents the increase in regions from 18 at the time of the 1967 census to 20 in 1978, 1988 and 2002, and to 25 in 2012. The data assembled below uses the 20 region structure.

A2. TzTownPops67-12

Sheet 1: 'Pops&Growth'. Here figures from A3 and A4 below are combined to produce as complete a series as possible of urban populations for the period 1967 to 2012. Towns with population over 10,000 are separately identified; the rest are grouped together as 'other urban'. In rows 318-328, summary figures for 1957 are added.

Sheet 2: 'Zipf plots'. Here towns with populations over 9,000 in 2002 are ranked for each of the years 1967, 1978, 1988 and 2002 and then log rank plotted against log population to obtain Zipf curves (see chart). The main indication from these plots is that the population of Dar is at least 4 times as big as would be expected.

Sheet 3: 'Growth charts'. Here charts are created to show (a) growth of regional capitals 1967-2012; and (b) growth of total population, rural population, urban population, Dar, regional capitals and other urban. It may be seen that by 2012, the urban population was roughly equally divided between Dar, regional capitals and other urban.

A3. TzUrbPops2012

Sheet 1: 'Best estimates'. Here the numbers obtained in Sheet 2 have been taken, and problems resolved as far as possible by checking against district urban populations, comparison with figures published by Th. Brinkhoff and direct queries to NBS. Remaining uncertainties are reflected in col H 'Notes/Queries'.

Sheet 2: 'from 2012 Census'. The construction of this table started from a list of the names of 'Urban Centres' recognised by NBS for the 2012 Census, together with their ward codes. These were then listed in strict code order and ward populations (total/rural/urban), areas and densities were then added. Urban centre populations (col F) were then taken from ward urban populations (col N). Problems encountered were dealt with as follows: (i) Where an urban centre contains many wards, the urban populations have been summed and cols H to P left blank; (ii) Some wards contain more than one urban centre (e.g. in Dodoma region, Mpwapwa ward includes Iloilo, Mpwapwa and Mwanakianga). In such cases, the individual urban centre populations are uncertain; (iii) In some cases (e.g. line 21, Arusha region, Kiranyi ward), the urban centre has a different name. In these cases, the ward name has been added in brackets; (iv) In some cases (e.g. line 48, Kilimanjaro region, Kisima ward), there is an urban population but no matching urban centre name. In these cases, the ward name is used in brackets; (v) In a few cases (e.g. line 294, Tabora region, Kaliua), there is an urban centre but no matching ward urban population, so the population has been put as '?'.

A4. Tanz UrbanPops 1967-2002

Sheet 1: 'Best 67-02'. The urban populations for 1967 are taken from the 1967 Population Census, Vol 2 'Statistics for Urban Areas' (Bureau of Statistics, Ministry of Economic Affairs and Development Planning, Dar es Salaam, 1970). Figures for other years are taken from sheets 2 to 4, as explained below. Cols A to H bring together urban populations from sheets 2 to 4 and the 1967 Census by region. It may be seen that some urban areas appear in one year but not in others. This underlines uncertainty about some aspects of

Tanzania's urban development, particularly as few of these smaller town names appear on any of the standard maps of Tanzania (a good research project for someone local?). Cols L to S then focus on urban areas with population over 9,000 in 2002, with the balance of the urban population in each region recorded as 'Other urban'. Cols U to Y then show growth rates for various periods.

Sheet 2: '2002 list'. Urban populations for 2002 are particularly difficult to pin down. Enumeration Areas were classified as 'urban' or 'rural'. The 20 regional reports on the Census go down to ward level and where a ward contains some urban EAs and some rural EAs, it is classified as 'mixed'. Cols C, D and E list all towns or wards with 'urban' or 'mixed' populations. It may be seen that there a very large number of 'mixed' entries. To try to get the correct urban populations, cols F and G give urban populations reported by Th. Brinkhoff (for 2002) and E-Geopolis (for 2000) respectively. For the 'Best' column H, the Census figure is taken if unambiguous; if not, the Th. Brinkhoff figure is accepted (it appears likely that Th. Brinkhoff had access to more detailed Census data); in the remaining cases, the E-Geopolis figure is used, increased by 10% for the timing difference. Doubtful figures are highlighted in yellow. Some names, marked 'W' are thought to be wards, not towns. For example, it was later possible to establish that Bagamoyo and Bulyaga wards in Mbeya region together make up Tukuyu town; and Kawajense and Shamwe wards in Rukwa region are part of Mpanda town. Cols L to T then focus on urban areas for which a 'best' estimate population has been found. These populations total 6.743 million (incl. some minor duplication), compared with the official census mainland urban population of 7.555 million, indicating some 0.8 million of unidentified urban populations. Part of the discrepancy is attributable to Arusha: the official urban population for Arusha MC is 333,791 but this is hard to reconcile with the Census report which suggests an urban population of around 260,000. Nevertheless, the official figure is used in Sheet 1, to avoid the difference showing up in 'Other urban'.

Sheet 3: '1988 list'. Despite the 1988 census giving little direct attention to urbanisation, it proved possible to find a large number of population figures for urban areas, as shown in col E by reference to the 20 Regional Profiles published as part of the Census reports. The main difficulty is knowing which names are wards and which towns. Names thought to be wards are marked 'W'. For comparison, the figures were compared with those published by Th. Brinkhoff (for 1988, col F) and by E-Geopolis (for 1990, col G). Also shown are urban population figures included in a table in the National Human Settlements Development Policy, published in 2000 (col H). However, for the most part, these last figures are considerably larger than those found in the census, for reasons which are unclear. In all but two cases, the census figure has been taken as 'Best' (col I). For Bomangombe/Hai the E-Geopolis figure less 10% for the difference in timing is taken; for Tukuyu, the sum of Bagamoyo and Bulyagi wards is taken. The total of the 'Best' column is then 4.015 million, compared with the census mainland urban population of 3.998 million, indicating about 17,000 of double-counting. In sheet 1, only towns with urban population over 9,000 in 2002 are used, the balance appearing as 'Other urban'.

Sheet 4: '1978 list'. Here the position is much more straightforward. Table 5 of the 1978 Population Census, 1978, Vol IV gives urban area populations. It appears that the Th. Brinkhoff figures are based on the same source. The E-Geopolis figures, which are for 1980, appear less reliable. As for 1988, only towns with urban population over 9,000 in 2002 are used in sheet 1, the balance appearing as 'Other urban'.

A5. Tz Urb&Mig Propensities.

Sheet 1: 'Refugee adjusted'. This sheet shows how the census regional totals of rural and urban populations have been used to calculate three propensities:

- i. **PROM**, the regional propensity for rural out-migration: This is the percentage of the expected rural population in a region that migrates either to the urban parts of the same region or to other regions (a negative value indicating a net inflow to the region's rural areas);
- ii. **PUIM**, the regional propensity for urban in-migration: This is the number of migrants to the region's urban areas expressed as a percentage of the expected urban population (a negative value indicating that some of the expected urban population left the region's urban areas);
- iii. **PRIM**, the regional propensity for in-migration, both rural and urban: This is the number of migrants coming into the region expressed as a percentage of the expected total population of the region, rural and urban (a negative value indicating a net outflow from the region).

To obtain these measures, it is assumed that the natural growth rate for all regions between the census years 1978, 1988, 2002 and 2012 is the national average rate for each period. Of course, this is unlikely to be quite right but it provides a benchmark – the 'expected population' – against which other movements can be assessed.

In an attempt to remove the distortion caused by refugees in Tanzania (mainly from Burundi), the rural populations of Kagera and Rukwa regions have been reduced by 100,000 each in 1988 and 2002, and the rural population of Kigoma region has been reduced by 400,000 in 2002 and 100,000 in 2012. While these adjustments are based on UNHCR and other reports, they are very approximate – better estimates would be desirable.

Sheet 2: 'Not adjusted'. As sheet 1 but without the refugee adjustments.

A6. Dar wards pops areas 2002

Sheet 1: 'Wards02-PopsAreasDens'. This lists in cols A to I the wards of Dar region in ward code order, giving total population, area (sq. km) and density (persons/sq. km). In cols L to T the same data is shown in descending density order, with sub-totals taken at > 5/Ha, >2.5/Ha and >1.5/Ha. It may be noted that over 90% of the population live in wards with density >5/Ha but these wards take up only about 30% of the Dar area. At the same time, the Dar region includes a number of wards with very low densities (<1.5/Ha), particularly in Temeke area, where about 1% of the regional population occupy 35% of the area.

Sheet 2: 'Streets02-PopsAreasDens'. Similar to Sheet 1 but areal units are now streets (mtaa).

A7. Dodoma wards pops areas 2002&2012

Sheet 1: 'Wards02-PopsAreasDens'. Similar to sheet 1 of A4.

Sheet 2: 'Streets02-PopsAreasDens'. Similar to Sheet 2 of A4.

Sheet 3: 'Wards12-PopsAreasDens'. As sheet 1 but for 2012.

Sheet 4: 'DOD UrbanGrowth 02-12'. Comparisons of urban growth rates in Dodoma region, using census data and ward density or street density cut-offs.

A8. Dodoma wards Shapefiles (7files, zipped)

These are the standard set of Shapefiles. They do not include any additional layers so when processed through GIS software simply deliver a map of Dodoma region showing ward boundaries.

A9. 2012 Census Regions Wards PopsRur&Urb (30 files, zipped – one for each region, incl. Zanzibar)

These files are extracted from the full Census. They give Total Population (Both/Male/Female), Rural Population (Both/Male/Female) and Urban Population (Both/Male/Female) for each region, district and ward.

B. Other

B1. Crop areas 81, 88, 02-3, 06-7 (2 sheets)

Sheet 1: 'All crops 81,88,06-7'. Lines 1-21 show regional figures for area planted in 1981 for cereals (Maize, sorghum, millets, paddy and wheat) and other crops (cassava, sweet potatoes, pulses, oilseeds & oilnuts). Lines 26-48 show the same data for 1988. Lines 52-74 are for 2006/7 and give just total area planted for cereals and other crops. The 2006/7 figures are from the Agricultural Sample Survey for small scale farms so may not be fully comparable with the earlier figures.

Sheet 2: 'Maize (AreaProdYield) 81-2_02-03'. The figures here cover all years from 1981/2 to 2002/3 and are taken from various editions of 'Agriculture: Facts and Figures' produced by NBS. Doubtful figures have been highlighted in yellow. (Note: Regions are listed in alphabetic order in this publication.)

B2. Road distances (2 sheets)

Sheet 1: 'PavedUnpaved 2003&2011'. This shows lengths of road by region divided into 'paved' and 'unpaved' from NBS publications. Also shown are figures for 'earth&gravel' roads in 1993, taken from a World Bank report, which are probably comparable with unpaved.

Sheet 2: 'Distances between regional caps'. Lines 1-21 show the reported distances between regional capitals. These have presumably changed little over time but the effective distances will be affected by road improvements over the years. As a step towards estimating 'market potential' for each regional capital, lines 23-43 propose 'distance discounts', which are then in lines 45-50 multiplied by regional urban populations and summed for each of the years 1978, 1988, 2002 and 2012. (The results do not show very big differences between regions: possibly larger distance discounts would be appropriate and/or poorly connected regions should be left out of the summation.)

B3. Palmer Drought Severity Index Tz (4 sheets)

The Palmer Drought Severity Index (PDSI) data for Tanzania shown here were kindly provided by Mathieu Couttenier, co-author of Couttenier M & R Soubeyran (2014) "Drought and Civil War in sub-Saharan Africa" *Economic Journal* Vol. 124, No. 575 pp. 201-244. As they explain: "We use the monthly grid cell data from Dai *et al* (2004). This database covers the world time series from 1870 to 2005; it is geolocalised and available at a resolution of $2.5^0 \times 2.5^0$ (about 250 km at the equator)." In the original, the index is scaled from -0.1 (<-0.04 = extremely wet) to +0.1 (>0.04 = extremely dry), with 0 as 'normal' but Couttenier & Soubeyran have re-scaled it to run from 0 to 30, with 15 = normal. In our use of the annual average data for Tanzania, we have taken values less than 12 to indicate an unusually wet year, and values over 18 to indicate an unusually dry year. The resolution gives eleven 2.5^0 squares in Tanzania and we have assigned regions to the square in which most of the region lies – see **Sheet 2: 'Raw data 1946-2005'** and **Sheet 4: 'regions 1961-2005'**. **Sheet 3: 'regional averages'** is derived from sheet 4 and shows average values for the periods 1961-67, 1967-78, 1978-88 and 1988-2002. **Sheet 1** compares the PDSI with some (unpublished) partial productivity indices for Tanzania agriculture kindly provided by Block.

Couttenier & Soubeyran (p.203) explain the advantages of the PDSI as:

"It is the most prominent meteorological drought index. This drought severity index is a function of the duration and magnitude of abnormal moisture deficiency. The PDSI captures meteorological conditions on the ground and combines contemporaneous and lagged values of temperature and rainfall data in a non-linear model (with thresholds). First, the index captures important interactions that were missing in previous studies. For instance, low rainfall is more important in hot months because evapotranspiration is significant and there is in turn less moisture recharge (or more loss if the layers are full). Indeed, high

temperatures can prevent abundant rainfall from recharging soil. Second, the index depends both on the limited capacity of moisture accumulation of the soil and on the local characteristics of the soil. As a consequence, abundant precipitation that reaches the accumulation capacity of the soil will run off (and will not be captured by the ground). Third, the PDSI takes the heterogeneity in local conditions and the differences in local climate history into account.”

Note: The PDSI measures conditions relative to a long term average for the region or country being observed so is not suitable to compare how drought-prone one region or country is compared to another. It does show whether a particular region or country is getting more or less drought-prone over time.

B4. Tz Employees & Earnings 67,78,88, 02 & 12 (7 sheets)

Up to 1978 regular surveys of Employees and Earnings were carried out. These cover formal sector employment only. During the 1980s and 1990s, there seem to have been no such surveys, presumably a reflection of the extreme pressure on public expenditure over most of this period. When the survey resumed in 2001, the regional figures were analysed by industry but not by urban and rural, as previously. In the 2012 survey, the industry analysis was replaced by a simple division into private and public; there was also an analysis between permanent and casual employees. In viewing the figures, it may be best to start with **Sheet 7: 'Summary'** which brings together the key national figures from the other sheets. It also has regional figures for 1969 and 1976 (as these happened to be the first I came across in SOAS library). Subsequently, the objective was to obtain figures for the census years 1967, 1978, 1988, 2002 and 2012 (and incomplete data for 1967 led to 1968 being used). **Sheets 1-6** are then largely self-explanatory. The data also allow average and relative wages by region to be calculated.