

The *DEMOGRAPHY-STATISTICS-INFORMATION TECHNOLOGY* Letter
FROM: Griffith Feeney feeney@gfeeney.com
Letter No. 1, 19 June 2013

The Population of Lagos

To analyze data we need to know where it came from, but what exactly does this mean?

It may mean a citation of the publication the data was taken from. But what if the publication is a secondary source?

Usually we want a citation of the primary source. Secondary sources may get it wrong. Often they omit important metadata. They may not indicate the primary source.

The fullest answer to “Where did this data come from?” is a description of the process that produced the data. This may require multiple documents running to hundreds of pages.

Knowing where data comes from is an essential practical discipline of data analysis. The following story shows why.

In 1994, when my family and I were living in Canberra, we were invited to dinner by Jack Caldwell and his wife Pat. Over dinner Jack told the following story.

He had read in the journal *West Africa* that the population of Lagos was five million persons. Having worked extensively in Nigeria, he was keenly interested in the number. Being a good demographer, he was equally interested in the source.

So he wrote to the journal asking for the source. They replied that the number had come from someone Jack knew at the University of Birmingham’s West African Centre, Margaret (Peg) Pell.

So Jack wrote to Peg, asking her for the source. She replied that she had heard the number from Bob Morgan, another friend. Jack guessed that the source was a demographic survey that Bob had worked on. And of course he wrote to to Bob for confirmation.

Bob replied that the survey was not the source, as it had not been completed, but the letter did provide the source.

You remember, Jack, that I picked you and Pat up at Lagos airport nine months ago. Your flight path had come in over the full length of the city and you remarked to me that it had grown greatly and now looked as if it might have five million inhabitants. I knew that you had flown over many cities and knew the populations of many of them, so I thought that this was probably the best estimate available. I have subsequently employed it when people have asked me the question.

The source that Jack had searched so assiduously for was—himself!

The story shows that to assess data quality we need to look not just at the data itself, but also at the process that produced the data. Data quality is a function of the nature of the data collection process and the care with which it was executed.

The estimate in this case is remarkably good, as the resources section below shows. Given the process, however, we would not expect high precision.

0.1 Resources

The United Nations Population Division produces estimates of the population of major cities of the world. *World Urbanization Prospects: The 2011 Revision* is available for free download via the “Final Report” link at esa.un.org/unpd/wup/.

The “Urban Agglomerations” link provides access to an online database. The estimates for Lagos for mid-1990 and mid-1995 are, respectively, 4.8 and 6.0 million.

DEMOGRAPHY, STATISTICS, and INFORMATION TECHNOLOGY is an every-other-month email letter from Griffith Feeney, Ph.D., an international consultant based in Scarsdale, New York, USA. For more information see griffithfeeneyconsulting.com.

This letter and past letters are available at demographer.com. To SUBSCRIBE, send email to feeney@gfeeney.com with “subscribe” in the subject field. To UNSUBSCRIBE, send email to the same address with “unsubscribe” in the subject field.