

A CASE STUDY OF THE INTERNATIONAL COLLECTION OF DEMOGRAPHIC STATISTICS

FORREST E. LINDER¹

EARLY in 1947 the United Nations Statistical Commission made a recommendation that "demographic data concerning all the countries of the world be assembled and published by the United Nations." Two weeks later this recommendation was made more specific by the Population Commission, which recommended the publication of a Demographic Year Book and that the "target date of publication of the first issue be 1948 and as early as practicable in that year."

Simultaneous with the recommendation that this task be undertaken by the United Nations Secretariat, the Population Commission formulated the further instruction that "statistical matter published in the year book shall be taken from the official statistics published by the separate governments, or supplied by the separate governments, or calculated by the Secretariat with the consent of the separate governments concerned."

The problems of the international collection of demographic statistics can be studied in many ways. For each variable of interest in the population or in the vital statistics field, there are the problems of geographic coverage, there are questions of quality and accuracy of the obtained data and there is the difficult subject of the comparability of the definitions and classifications used by each country in its own collection and compilation. A systematic investigation and discussion of these questions would be a life-time task. And rather than to present a fragment of such a survey, in these few minutes, I have thought that it would be of greater interest to give a brief factual analysis of the results to date, of the above recommendations of the Statistical Commission and the Population Commission. In contrast to a general or theoretical discussion,

¹ Chief, Population and Vital Statistics Section, Statistical Office, United Nations.

this is more in the nature of a single clinical case history. Even this case history is further limited to a discussion of geographic coverage, without considering the equally important and much more difficult aspect of the quality of data.

The recommendations of the Commissions called for a broad coverage of data both geographically and by subject. At the same time the restrictions as to the sources of data to be used were rather rigid. These two considerations practically determined the general procedure for collection. The present library facilities of the United Nations do not include a complete collection of official statistical publications. The use of secondary sources is almost excluded by the limitation that data published must be officially supplied data.

Under these circumstances it was decided to send to each country a detailed questionnaire requesting *de novo* all of the historical and detailed demographic data which would be used in the year book or which, entered in the files of the Statistical Office, would form the foundation for future compilations or special studies.

A questionnaire was prepared containing thirty-three tables giving time trends and cross-classifications of area, population, births, deaths, stillbirths, marriages, and divorces. In addition to statistical data, detailed statements were requested on definitions, accuracy of data, special groups included or excluded from the figures, and other information which would be important for the correct interpretation of the statistical figures.

This questionnaire was prepared and printed in three languages—English, French, Spanish. Concurrent with the preparation, translation and printing of the questionnaire, consideration was given to the list of countries or areas to which the questionnaire would be sent. A geographic survey of the various political units of the world resulted in a list of approximately 250 sovereign countries, dependencies, colonies, trust territories, condominiums, international administrations, etc. This list of 250 areas became the basic distribution list for the demographic questionnaire, and represented the administra-

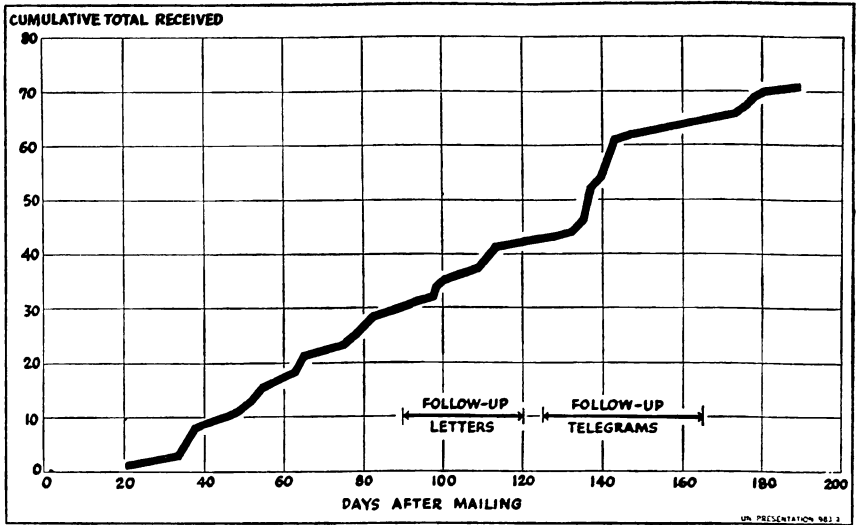


Fig. 1. Cumulative total of U. N. demographic questionnaires received, October 15, 1948.

tive sub-divisions of the world for which an attempt would be made to obtain demographic information.

The English language questionnaires were mailed in late March of 1948, the French questionnaires the middle of April, and the Spanish questionnaires early in May. All were sent by air mail. Figure 1 shows the rate of return. Approximately three to four months after the original mailing a follow-up letter was sent. During the fifth and sixth month a cable urging a prompt reply was sent to countries that had not answered. The chart gives some indication of the stimulating effect of the follow-up correspondence.

The net result, up to October 15, 1948, of these three formal contacts extending over a period of six months, with the responsible statistical agencies has resulted in the return of seventy-one questionnaires. At first glance it may seem that the international collection of demographic statistics is a very slow process. Yet the returns from this survey, taking advantage of the prestige and good will of the United Nations, are in general far better than has been the previous experience of international agencies. Slow as the process is, these returns

represent a high degree of cooperation on the part of the national statistical offices.

Figure 2 gives more specific information on the geographic distribution of the countries which returned or did not return a questionnaire. The seventy-one returned questionnaires do not by any means all represent important areas of the world. They range, in population, all the way from Campbell Island with a total population of five lonely male souls, to China with over 400,000,000. All together, the seventy-one returns cover 66.4 per cent of the world's population. And the approximately 180 unreturned questionnaires cover the remaining third. An examination of the map indicates that we have failed primarily to get returns from most of Africa, from the majority of the countries of the Middle East, from the countries and dependencies of southeast Asia and from the U.S.S.R.

It is the main purpose of this paper to discuss the adequacy of the data given in the seventy-one questionnaires which were returned up to October 15, 1948. It must not be assumed, however, that no data whatever exist for those areas for which a questionnaire was not returned.

To place the analysis of the returned data in proper perspective, a further examination is desirable for the areas for which data were not received. This examination may be made by considering separately the section of the bar chart in Figure 2 which represents the 33 per cent of the world's population for which no questionnaire was received. Figure 3 shows this part of the bar divided into sections according to the probable status of data for that portion of the world's population.

Speaking in general terms, it may be said that practically no data of any kind exist for about 8 per cent of the population. This segment includes such demographically unmeasured areas as Ethiopia, Afghanistan, Haiti, and most of the countries of the Middle East. For these areas, systematic data, either of the census or of the registration type are practically non-existent.

Another 8 per cent of the world's population can be put in

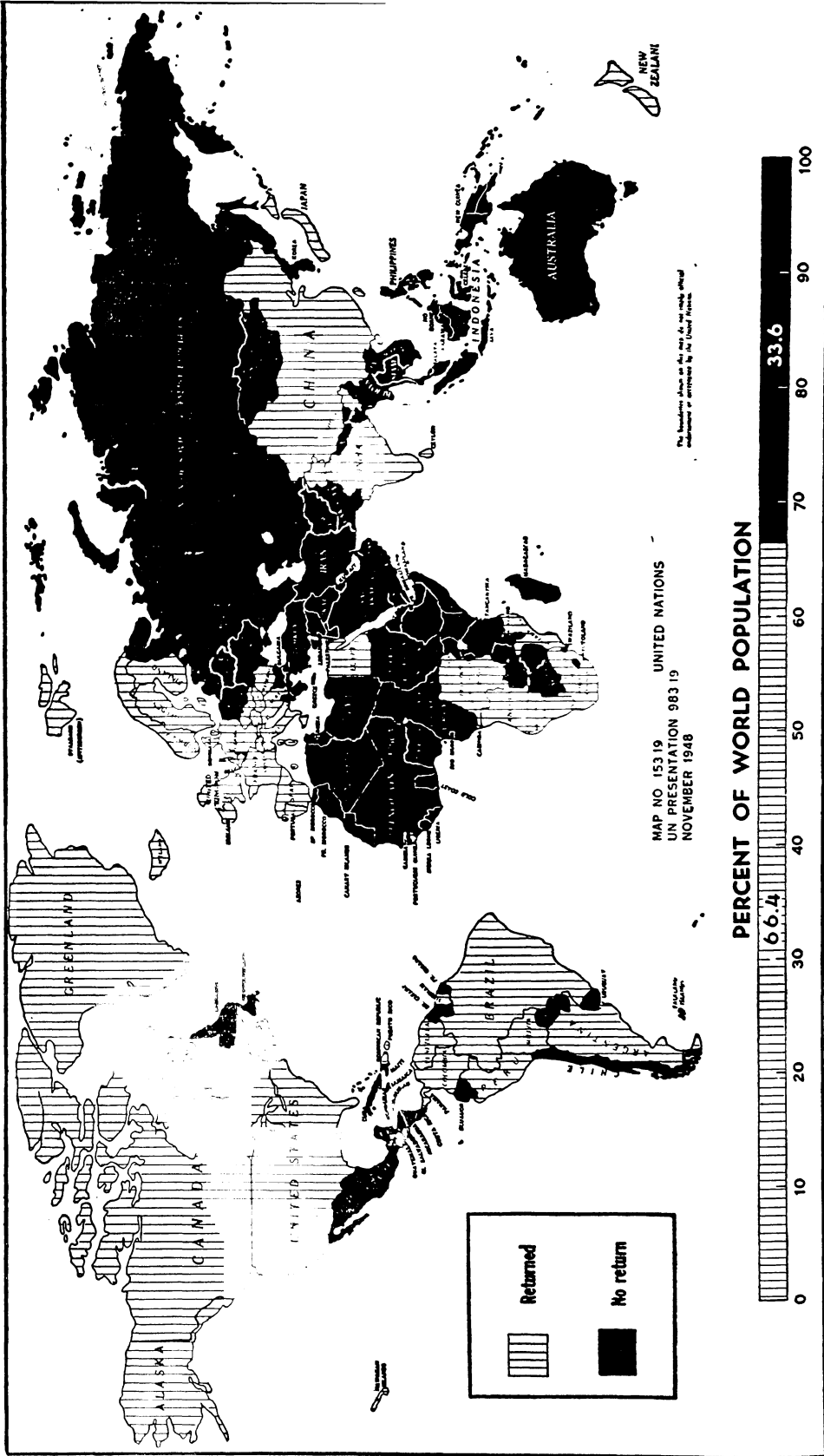


Fig. 2. Analysis of returns from U. N. demographic questionnaire, October, 15, 1948.

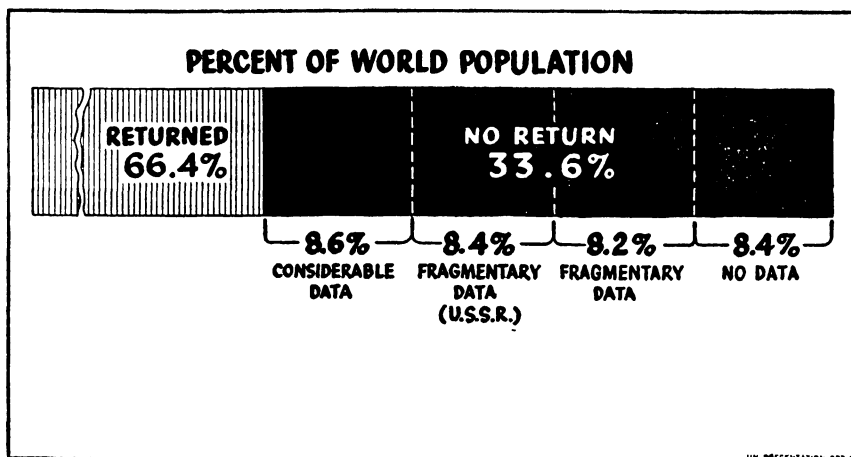


Fig. 3. Data for populations *not* included in U. N. demographic questionnaire returns.

the group for which fragmentary data only are available. For these areas there may be a summary census which gives total population with some data on composition. For other areas in this group there may be available some registration data but no population enumeration. This group of countries includes practically all of the colonies, dependencies and trust territories of Africa, most of the Caribbean and some countries of South America, and many parts of Southeast Asia.

Another 8 per cent of the world's population for which no questionnaire was returned is the population of U.S.S.R. This area may also be included in the group for which only fragmentary data is available. From the 1939 U.S.S.R. population census, total population and some classification by characteristics can be obtained. But other than this, little uniform systematic data of recent date is available. Demographic analyses of such areas can be made only by a laborious process of fitting together the available fragments and supplementing these with estimates and speculation.

The remaining proportion represents parts of the world for which fairly adequate data presumably exist but for which this particular questionnaire survey up to October 15 failed to get returns. This group includes such areas as Australia, Bulgaria,

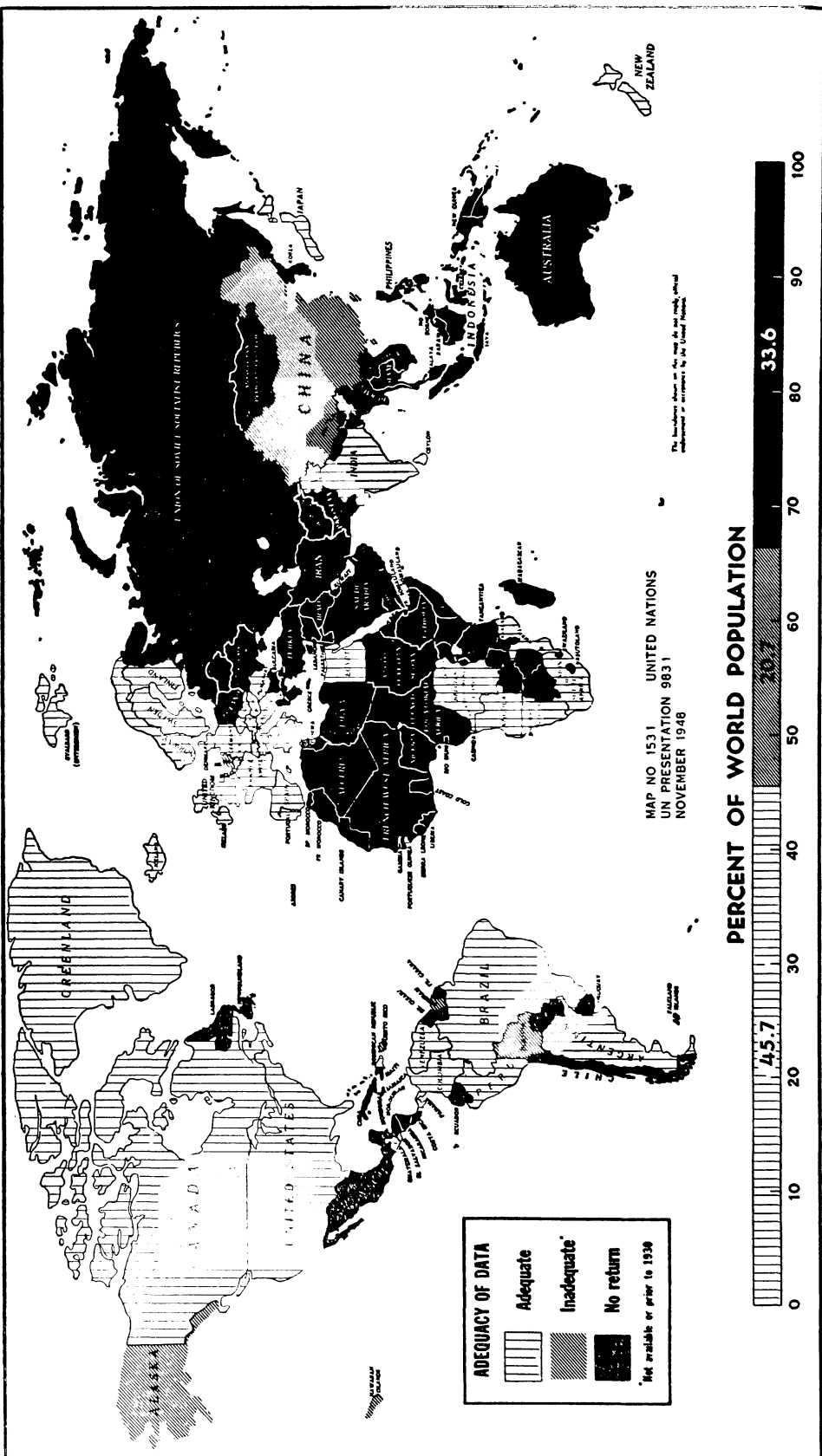


Fig. 4. Enumerated population. Analysis of returns from U. N. demographic questionnaire, October 15, 1948.

Chile, Greece, Mexico, and Poland. For some of these areas a substantial amount of data is available. For certain of them, we are informed that questionnaires are being prepared or are in the mails. For others, pre-war data can be found but disturbances of the recent war have not permitted the establishment of adequate post-war statistical services. Potentially, this final 8.6 per cent could be added to the 66.4 per cent for which data were returned.

The above discussion may be summarized in this manner. Questionnaires were received for approximately 66 per cent of the world's population. If this analysis were postponed for several months and if questionnaire data were supplemented by available material from library sources, this 66 per cent could be increased to 75 per cent. For approximately 17 per cent of the population some fragmentary data could be found and for the remaining 8 per cent essentially nothing is available. In examining the data for areas for which a questionnaire was returned, the probable status of data for the other areas of the world can be kept in mind.

A summary analysis of the data returned on the questionnaires can be made by examining a few typical items. Figure 4 shows the countries which answered the questionnaire giving an enumerated census figure for the total population. In this and the following charts, three broad classifications of the data returned on the questionnaires can be used. The first classification includes the approximately one-third of the world's population for which no questionnaire was received. The second class may be termed "inadequate data." This class refers to areas for which no information was given or for which the information related to a period prior to 1930. If no information for a given item is available since 1929, the status of data for that item for that area is defined for the purposes of this study as inadequate. The third group, "adequate" data, includes areas for which some information has been given on the questionnaire for some year since 1929. Admittedly this definition of "adequate" is rather liberal, but a definition referring to a

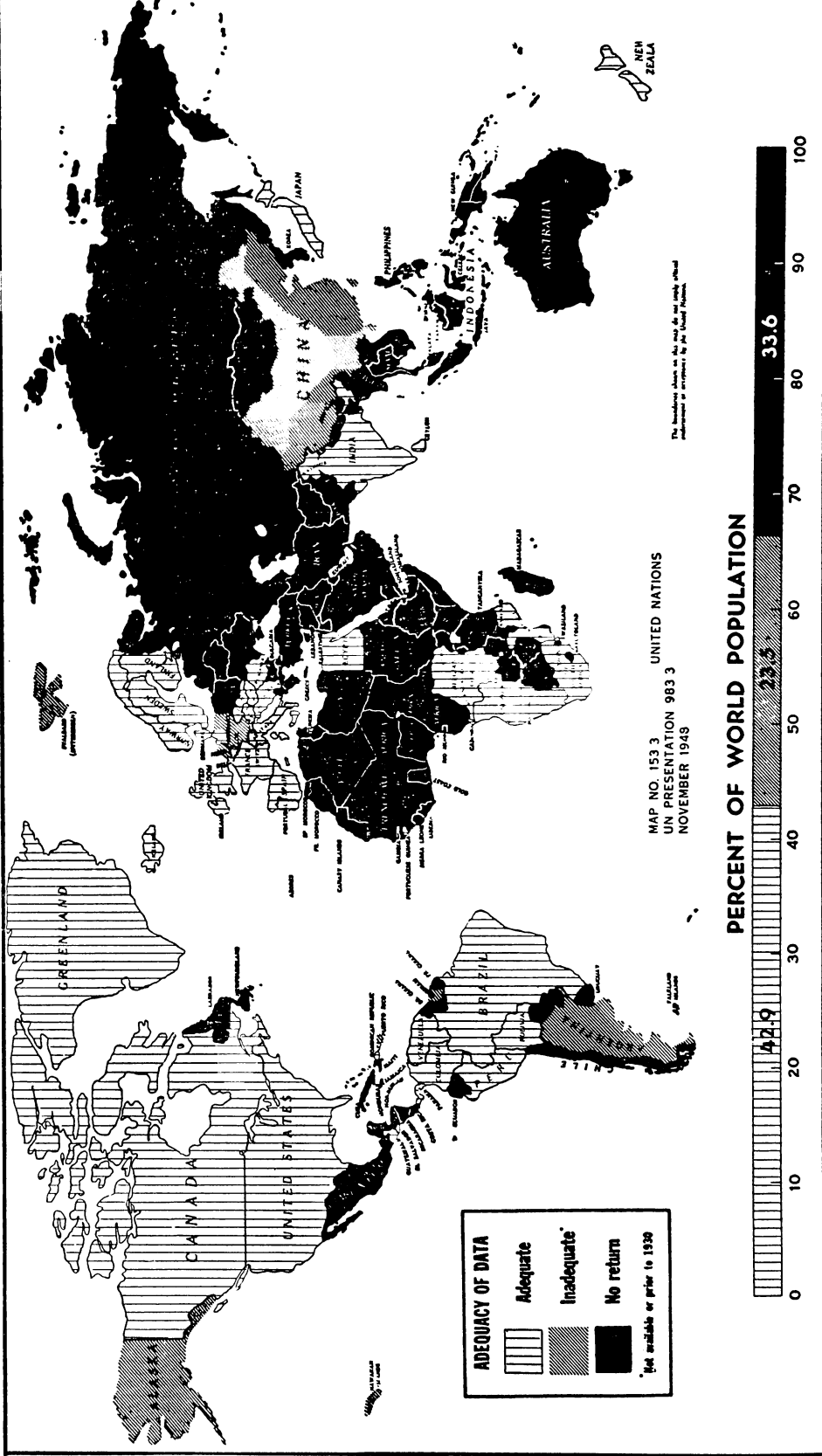


Fig. 5. Population by age and sex. Analysis of returns from U. N. demographic questionnaire, October 15, 1948.

shorter time period would have eliminated from the "adequate" group many highly statistically advanced countries whose regular decennial census tradition was interrupted in 1940 and 1941 by the war. It might be emphasized that the term "adequate" as used here, refers only to the simple existence of data, and not to the extremely complex question of the accuracy or international comparability of those data.

The shading on the map shows the various countries in each of these classifications and the bar chart at the base of the chart shows the per cent of the world's population represented by total population of the group of areas in each class.

Using these broad categories, a few charts will indicate the extent to which the questionnaire has been successful in collecting data.

As stated, Figure 4 shows the areas for which the questionnaire returns gave an enumerated population total. From this chart we see that for approximately 45 per cent of the world's population the questionnaires obtained a census figure pertaining to a year since 1929. Practically all of the returned questionnaires gave this information with the major exception of China which makes up almost all of the 20 per cent marked "inadequate."

In reference to this item of enumerated population, it may be stated that the greater part of the "no returns" group could be classed with the "adequate." From other sources it is known that about 75 per cent of the world's population has been covered by a "census" since 1929. It does not follow that all of these "censuses" have met high standards of a modern type enumeration, but at least for population totals, they must be considered as in the census class.

Figure 5, which gives information on population by age and sex, shows essentially the same picture. Here it is indicated that age-sex data has been received for almost 43 per cent of the world's population. China again represents the major unit for which a questionnaire was received but for which data on this item are lacking. This is also true for all following popu-

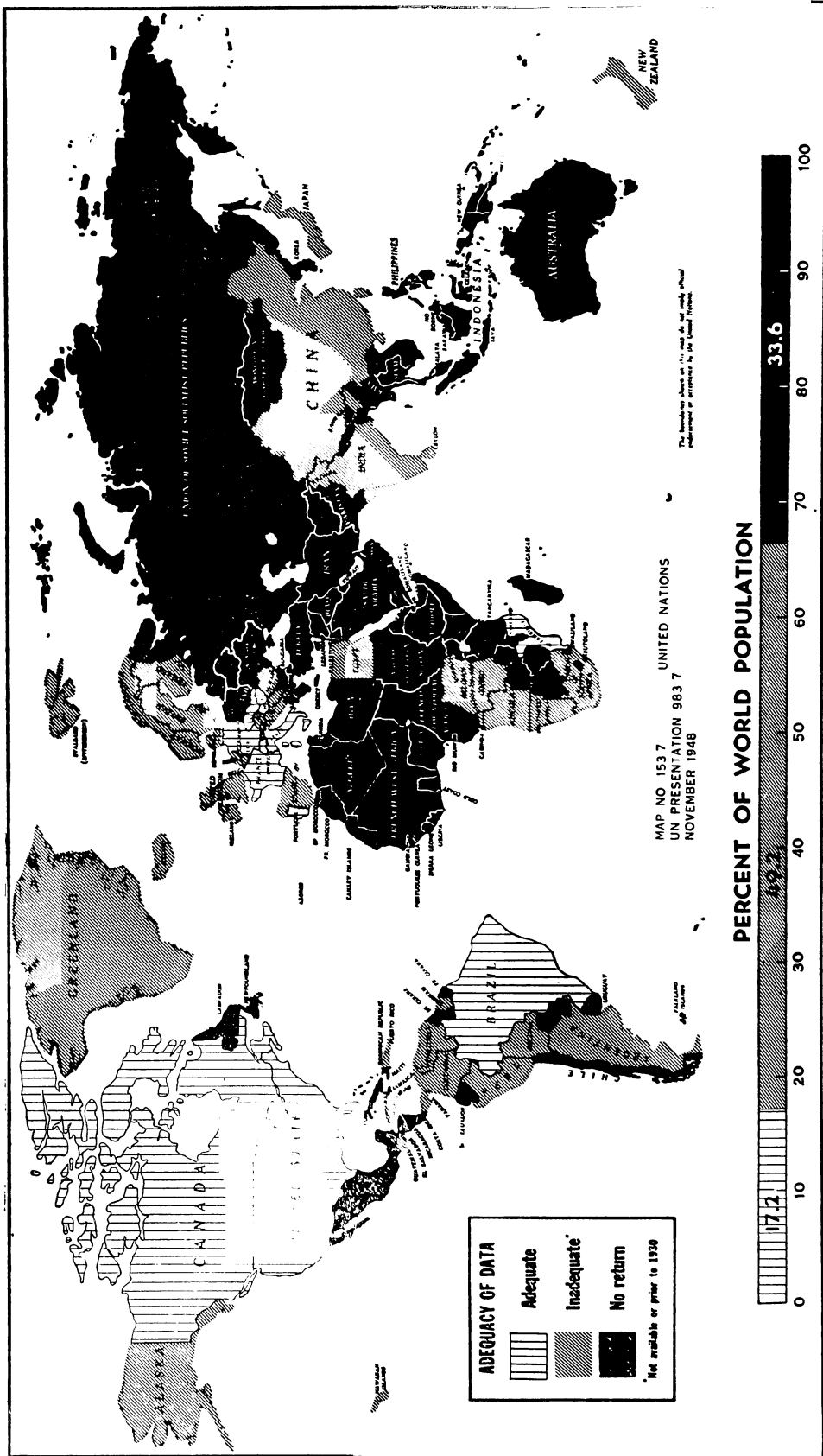


Fig. 6. Female population by number of children ever born. Analysis of returns from U. N. demographic questionnaire, October 15, 1948.

lation and vital statistics items. In general if a census includes more than a simple count of persons, it includes age-sex information. For this reason a good proportion of the areas for which no returns were received probably could be ultimately included in the group for which some data are available.

If we were to examine maps for other standard census items such as marital status, urban-rural distribution, the charts would be very similar to this one, although even more serious questions regarding comparability of definitions would arise. However, in reference to another group of population items, there would be found to be a much less uniform coverage. Such items as race, citizenship, language, country of birth are not of universal interest and each appears as a census item only in those countries where it is of particular importance. This type of item appears on most lists of recommended census items, but their interpretation and application differs from country to country. They may be considered as demographic items of particular, not of general, interest and should be collected only for those areas where they are sufficiently valuable.

Other items, however, of almost completely universal applicability from the standpoint of demographic analysis are even less available. Figure 6 shows the returns on the questionnaire for the distribution of female population by number of children ever born. Such data were returned for only 17 per cent of the world's population. Previously, in reference to total population and age-sex data, it will be recalled that a large part of the 33 per cent included in the "no returns" group were considered as potentially in the adequate group. However, in reference to items such as this, the questionnaire returns probably include practically all of the data which actually exist. The extremely inadequate information on fertility data of this type is further emphasized by Figure 7, which shows the areas giving data on female population by number of children now living. Such information was received for only 2.5 per cent of the world's population and it is doubtful if much of the 33 per cent of the "no returns" could be moved into the

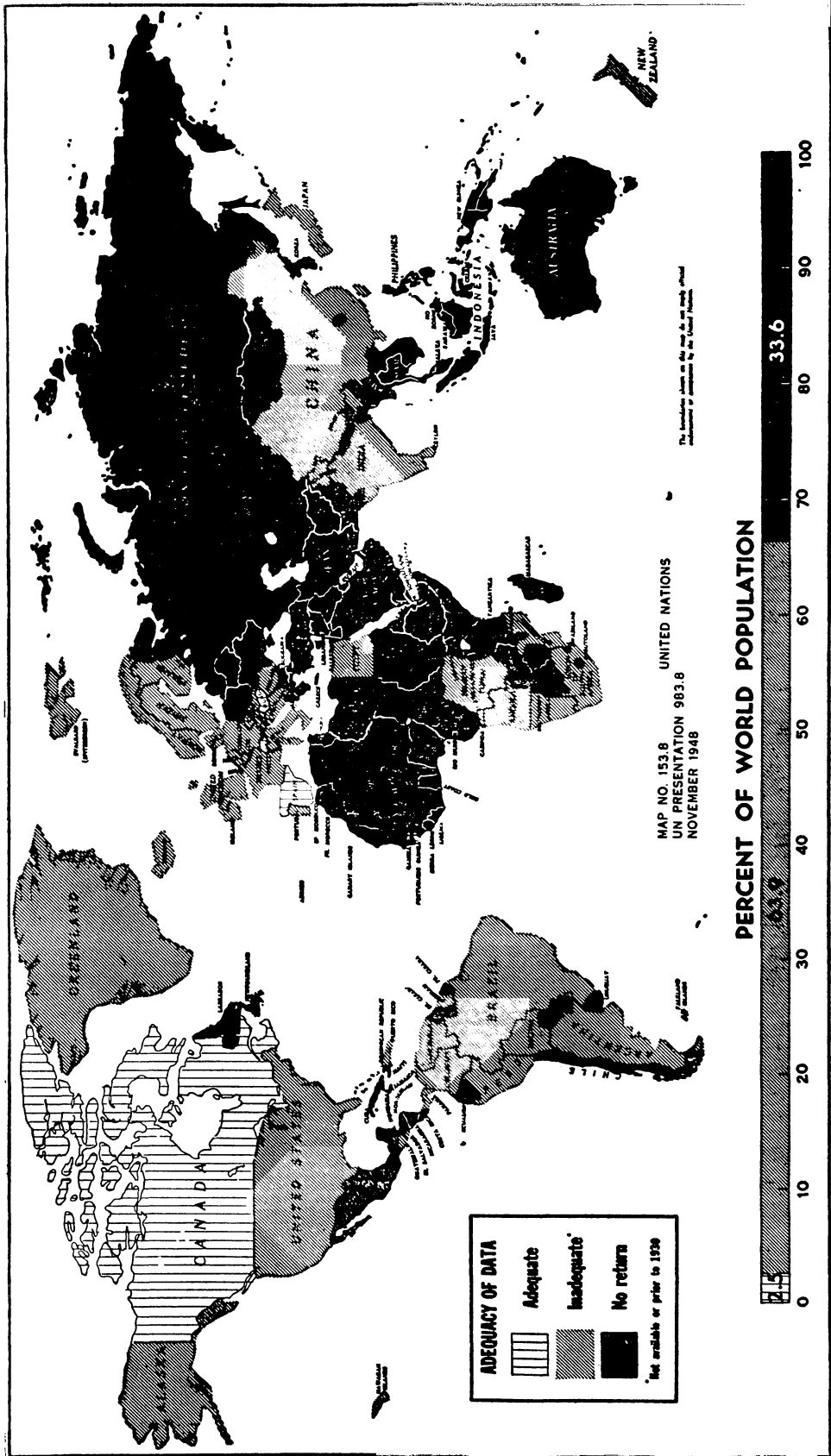


Fig. 7. Female population by number of children living. Analysis of returns from U. N. demographic questionnaire, October 15, 1948.

“adequate” classification. It is obvious that demographic analyses derived from material of this kind can have only a very limited geographic interpretation.

The adequacy of vital statistics, that is registration type data, is in general somewhat less satisfactory than for enumeration data. In making a dichotomy between “adequate” and “inadequate” data, the same criterion has been used—namely the return on the questionnaire of some data for a given item for some year since 1929. However, if vital statistics are available for any one year in this period they are usually available for most years except recent ones, because registration is a routine annual procedure rather than a periodic procedure such as an enumeration.

Figure 8 shows the areas for which birth and death totals were returned on the demographic questionnaire. It may be seen that such total data are available for only approximately 40 per cent of the world’s population. This percentage could be increased only by the addition of a very small amount from the 33 per cent of the “no returns.” Large areas such as the U.S.S.R. for which some census type data can be found, have made available almost nothing in the field of vital statistics. Similarly, the African colonies, Southeast Asia, and other areas which sometime during the past eighteen years have compiled some data on total population, have produced only fragments of registration data.

When it is stated that total birth and death figures are returned on the questionnaire for approximately 40 per cent of the world’s population, it must be repeated again that this figure is based on a liberal definition of adequacy. If the data were analyzed further to determine which data were acceptable from a standpoint of completeness and accuracy of registration, the percentage of the population for which good data are available would undoubtedly be much smaller. Even parts of the United States might fail to satisfy a reasonably strict criterion. Such an analysis must at some time be done, but at the present time few national statistical offices, and certainly

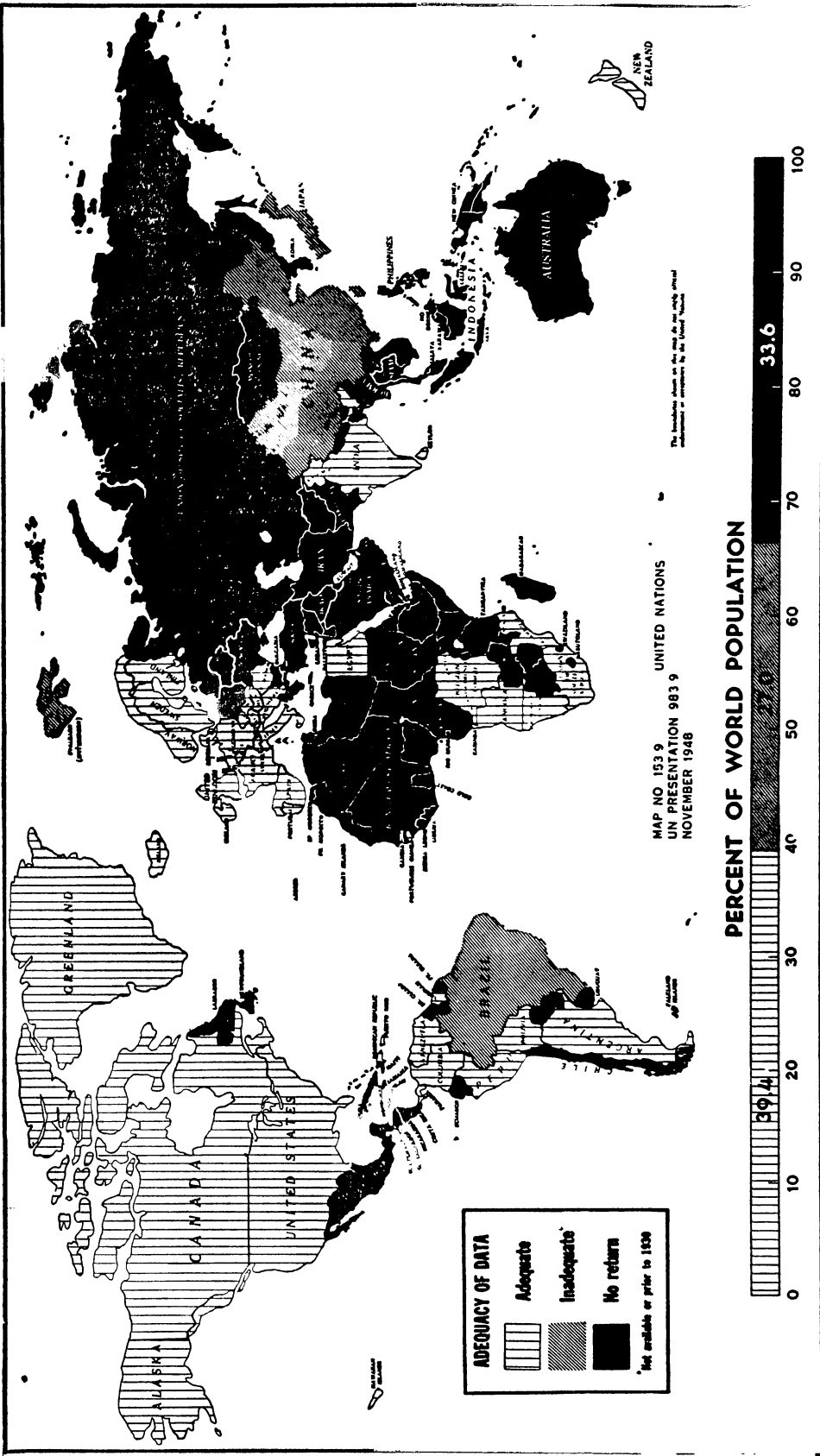


Fig. 8. Total deaths and live births. Analysis of returns from U. N. demographic questionnaire, October 15, 1948.

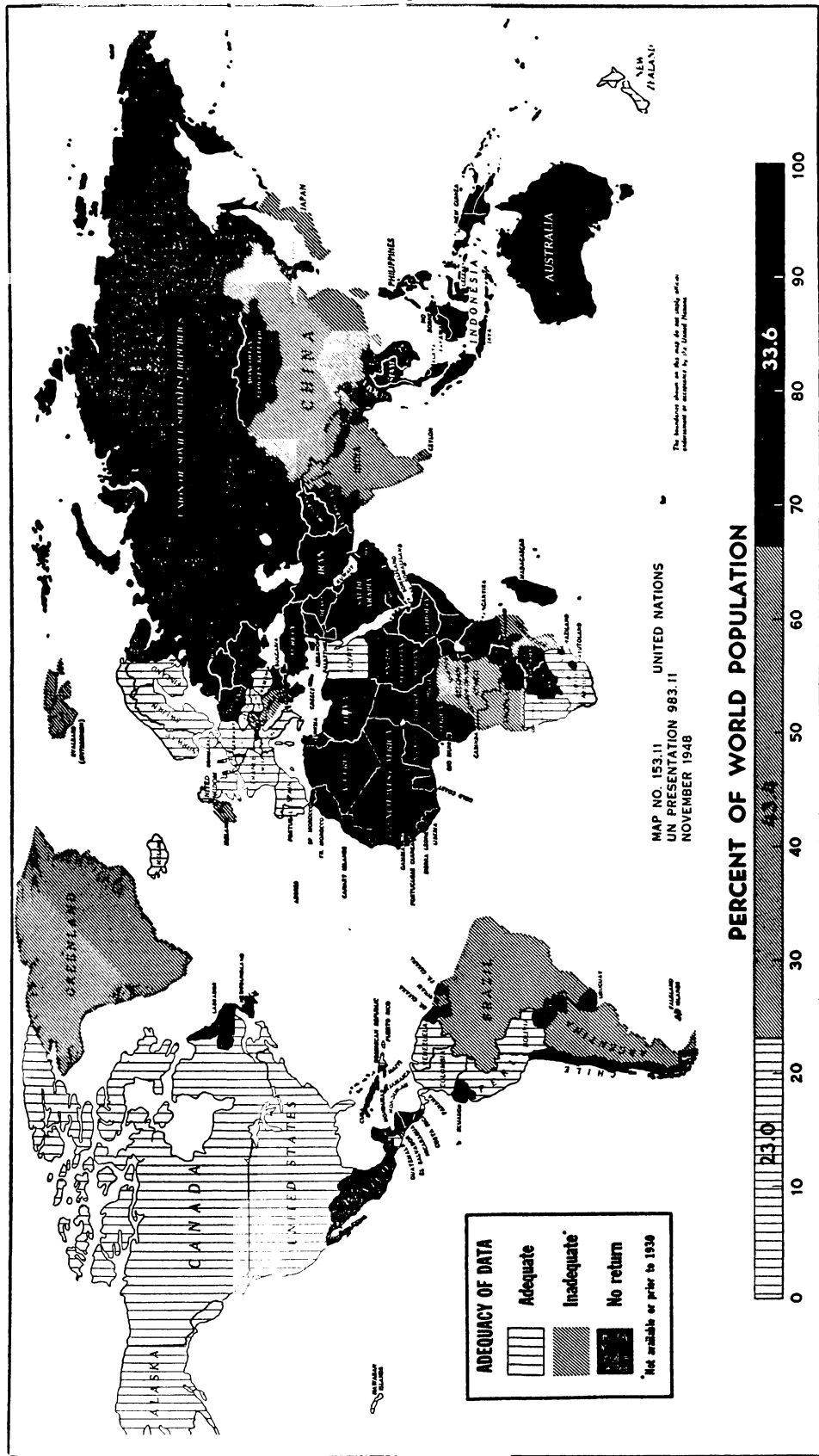


Fig. 9. Births by age of mother. Analysis of returns from U. N. demographic questionnaire, October 15, 1948.

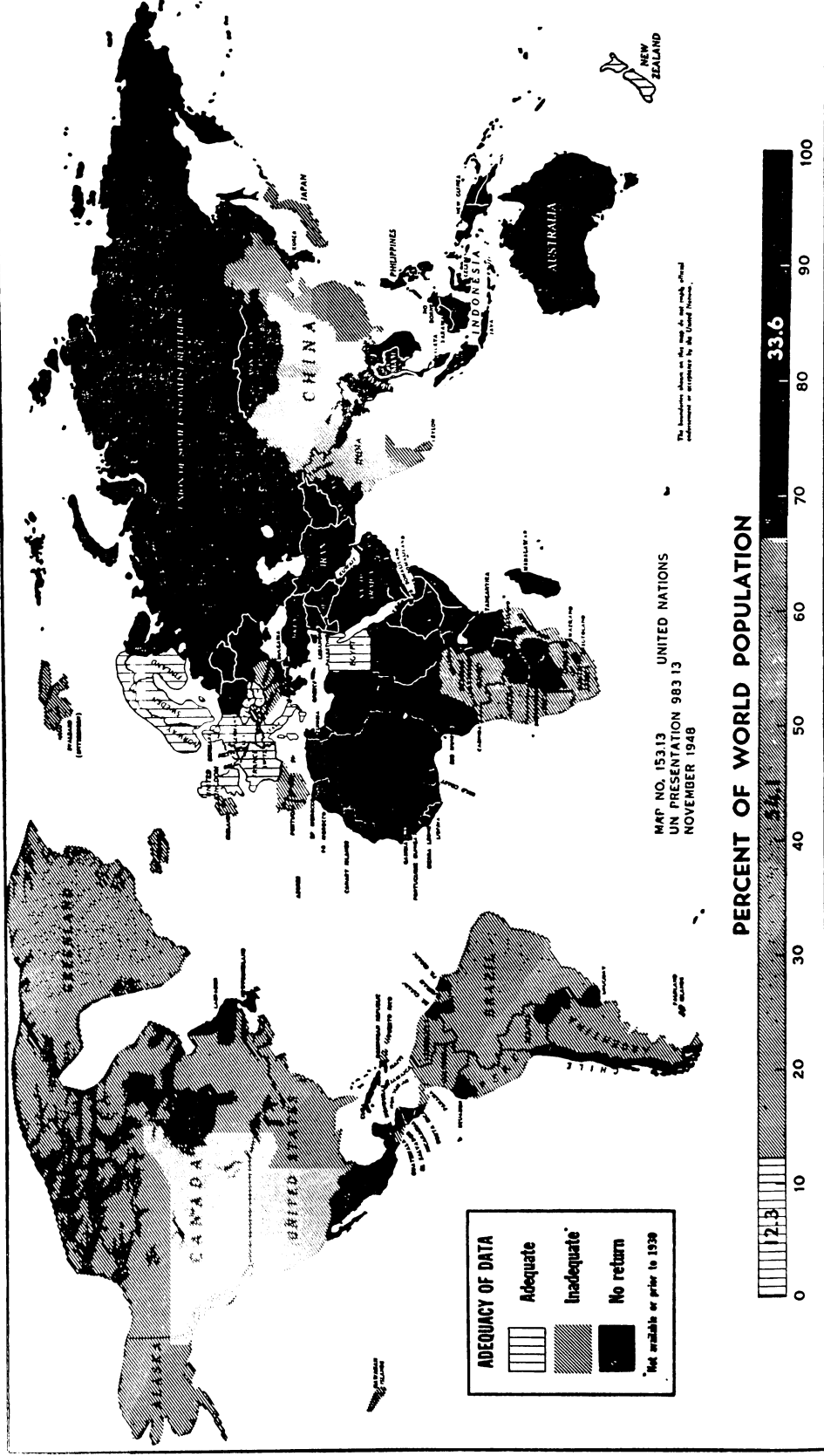


Fig. 10. Births by duration of marriage. Analysis of returns from U. N. demographic questionnaire, October 15, 1948.

no international statistical office, has the facilities or the facts necessary for an objective and thorough evaluation of quality of the data which it collects and publishes.

Adequate demographic studies cannot be based solely on total population, birth and death figures. Figure 9 shows the returns of data for births by age of mother. Information on this important subject was returned for only 23 per cent of the world's population. Large areas and sections of the world show as black or shaded areas. Here again it is doubtful that much of the dark areas can be transformed to hatched. And also for some of the areas shown hatched the available figures may not cover all of the population of that area. As can be seen, data by age of mother is available primarily for North America and Europe. Information on the number of births by the age of the mother and parity is available for the even smaller percentage of 19.

Figure 10 shows that data on number of births by duration of marriage have been returned for only 12 per cent of the world population. In this case the 144,000,000 population of the United States and the 12,000,000 population of Canada for which data are available on most all other items are missing.

The situation with regard to mortality data is slightly better. As seen previously, mortality totals were returned on the questionnaires for some 40 per cent of the total peoples of the world. Breakdowns of this total are more generally given than for sub-classifications of birth data. Whereas births by age of mother was given for 23 per cent, deaths by age of decedent was returned for 37 per cent. The areas included are shown in Figure 11. Other classifications of mortality data, such as deaths by month of death, infant deaths by subdivisions of the first year of life, are given for approximately the same areas.

Of all the various types of general data with which a demographer is primarily concerned, data relating to marriages seem to be the most inadequate. Figure 12 shows the areas which returned information on total marriages. These include scarcely more than one fourth of the world's population. Marriage data

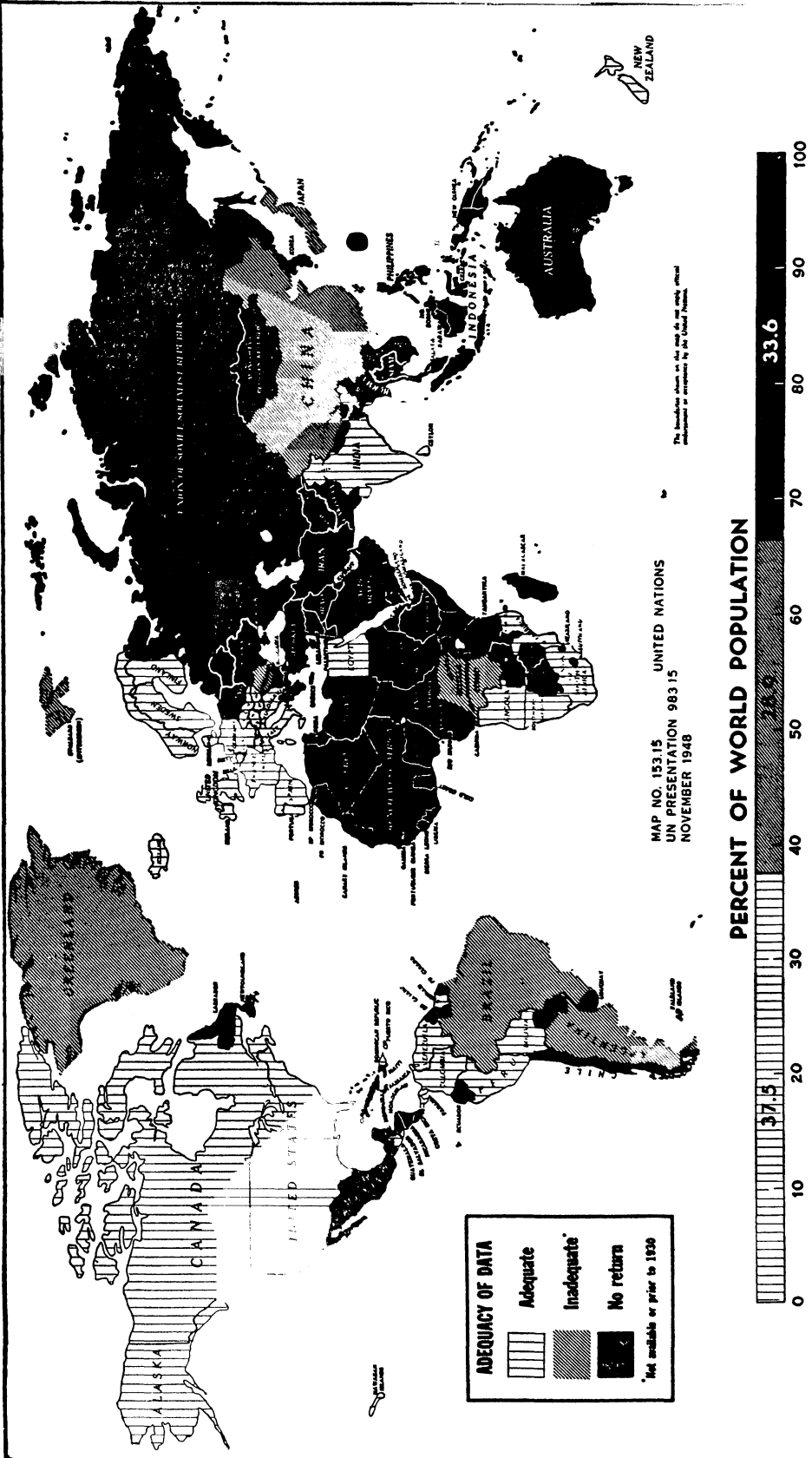


Fig. 11. Deaths by age and sex. Analysis of returns from U. N. demographic questionnaire, October 15, 1948.

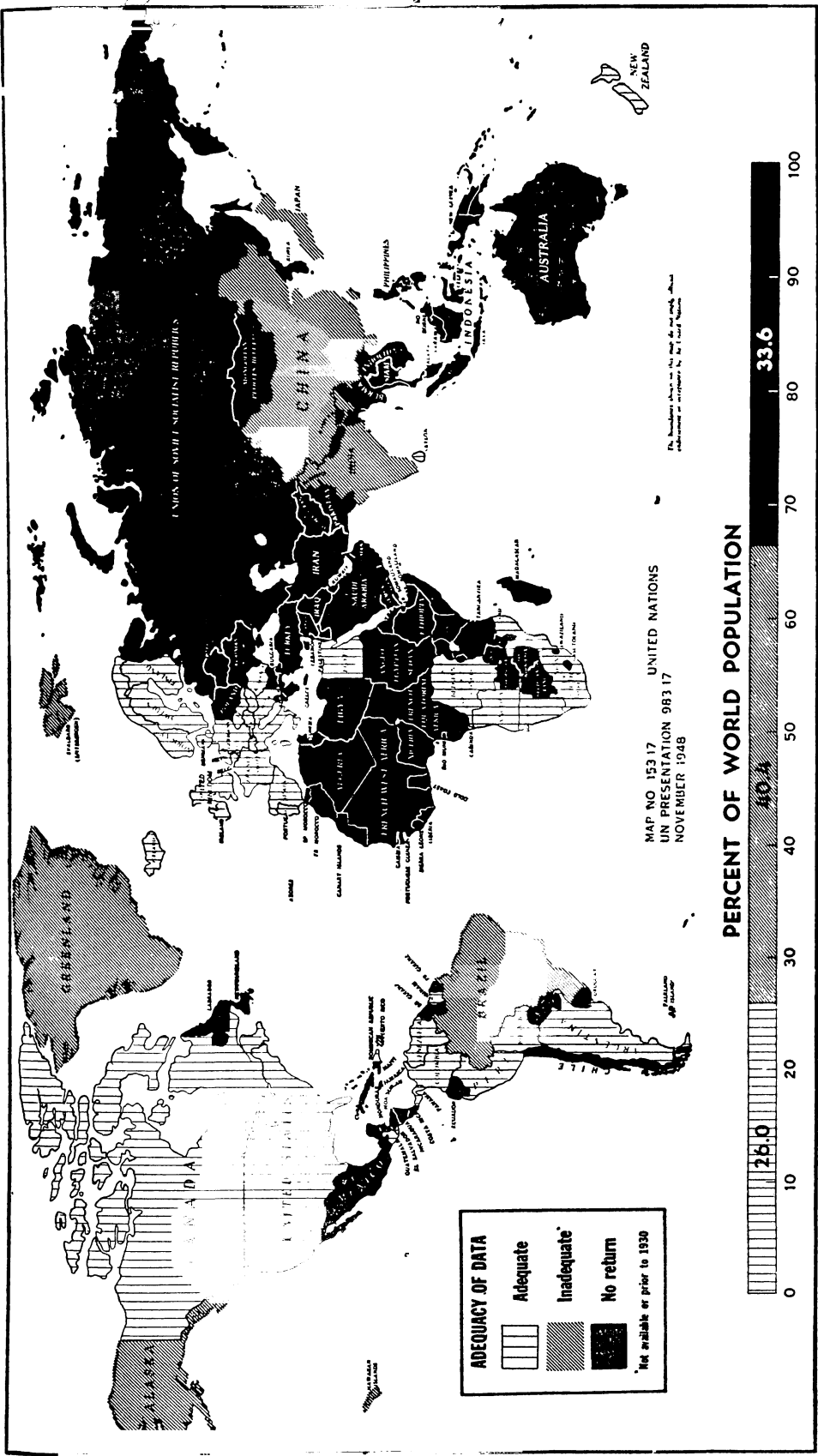


Fig. 12. Total marriages. Analysis of returns from U. N. demographic questionnaire, October 15, 1948.

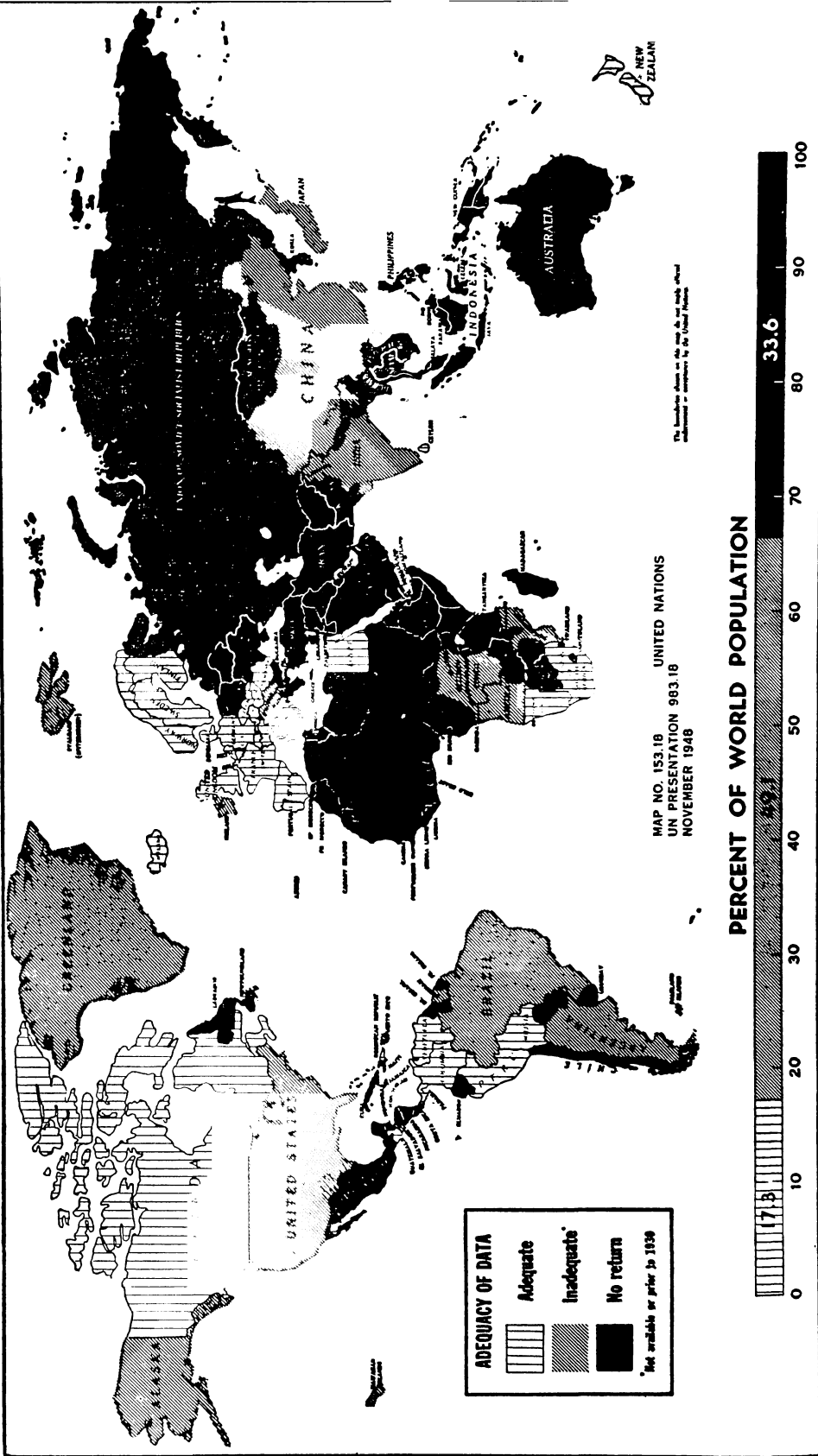


Fig. 13. Marriage by age of bride and groom. Analysis of returns from U. N. demographic questionnaire, October 15, 1948.

by characteristics, such as marriages by age of participants, are shown in Figure 13 to be returned for only 17 per cent. Looking upon the matter from this world-wide geographical viewpoint one conclusion regarding the United States becomes strikingly evident. That is, that marriage statistics—marriage statistics in the population census, marriage data in birth and death statistics, and statistics on marriage itself—have been very seriously neglected by the United States.

Perhaps if we can speak in very rough approximations, the facts shown on the various maps, together with an appraisal of the status of data for areas not returning a questionnaire, can be summarized in this manner. For approximately 75 per cent of the world's population it is possible to get a fairly recent (since 1929) enumerated or total census figure. The accuracy and detail for these figures would vary enormously. For the most usual breakdowns of population data, i.e. by age, sex, and marital status, data can be found for possibly 50 per cent of the world's total number of people. But for items more specially related to fertility analysis, such as female population by number of children ever born, certainly less than one-fifth of the world is covered.

Total birth and death figures are available only for slightly more than 40 per cent of the world population, and detailed classifications of birth figures by fertility items can be obtained for about one-fourth of the world.

These estimates can be highly discouraging or highly encouraging depending upon the philosophical point of view which you wish to take. It is certainly obvious that the data which are available are in no sense a random or a representative sample of the populations which we wish to study. And, accordingly, it must be concluded that it is not possible at this time to make a scientific and adequate presentation of the existing demographic situation or of trends for the world as a whole. This deficiency is unfortunate and may be even catastrophic because in a closely interrelated world a neat demographic analysis of one region for which data may be available is not

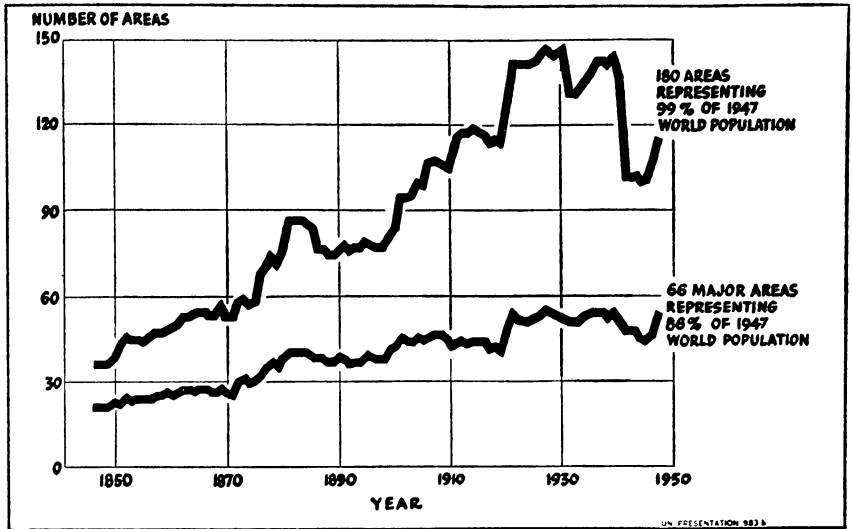


Fig. 14. Number of areas having a "census" within a decade of each year, 1847-1947.

necessarily an indication of what may be happening in the larger portion of the world which remains black on our maps.

From another point of view these estimates of available data can be viewed optimistically as a partial result of a long, slow but ever improving development of demographic statistics. The next chart will illustrate this point. Figure 14 presents certain facts from a compilation of data on the census history over the past century of some 180 areas of the world.

In 1947 these areas represented 99 per cent of the world's population. The line chart shows for each year of the past century the number of these 180 areas which, at any given year, had had a census within the preceding decade. The chart also gives a similar curve for sixty-six major areas representing 86 per cent of the 1947 world population. The downward drop at the end of the curves shows the disrupting effect of the war and this it is hoped will be compensated for by the unusual census activity focusing on 1950. The inevitable upward trend of these curves is an adequate demonstration that, if one is prepared to think in terms of generations or centuries, then demographic statistics are making substantial progress forward.

However and unfortunately, we may not be left alone to think in terms of generations and centuries. The demographic problems of the world may be upon us in much more urgent form before this slow process of statistical evolution has matured.

In this connection we may point out that there is a sharp difference between demographic statistics and some other types of statistics which the world uses to appraise its problems. In the fields of financial, trade, and production statistics, there is a strong positive relation between the existence of the problems and the existence of the data needed to study those problems. Countries with important foreign trade tend also to have foreign trade statistics. Countries with major financial responsibilities tend also to have the corresponding data. Countries with big production of oil, steel, or automobiles tend to parallel the production of material with the production of data measuring that activity.

However, in the demographic field quite the contrary is true. In this sphere it is found that countries with the greatest demographic problems are precisely those with the least data to measure and study those problems. Areas with the greatest relative production of people (in terms of births) and the greatest relative consumption of people (in terms of deaths) are exactly those for which human production and consumption figures are least available.

This negative correlation in relation to demographic statistics should indicate to the demographer that international compilations of official data will hardly serve his needs for studying his major problems.

The official agencies can gather in, collate, and publish material which is available for parts of the world where adequate census or registration procedures are functioning. But official agencies are not in a position to make a major contribution to the collection of data in the rest of the world.

It is obvious that for these other and major areas, new techniques and new methods must be found. Here is an open field

for the private investigator, the research worker who can construct a reasonable appraisal from the fragmentary data which can be found, or supplement these data with field studies and surveys.

It may be that substantial thought and resources should be directed to devising methods for penetrating the fog of demographic ignorance which envelops those areas outside the scope of official international statistical compilations.